# Peopling of the Americas: ABO Blood Group Haplotypes as an Indicator of Native American Origins and Migration from Siberia

By

Jacob T Boyd

Submitted to the graduate degree program in Anthropology and the
Graduate Faculty of the University of Kansas in partial fulfillment of the requirements
for the degree of Master of Arts.

_____

Chairperson: Michael H Crawford, PhD

_____

James H Mielke, PhD

_____

Bartholomew Dean, PhD

Date Defended: July 09, 2014

The Thesis Committee for Jacob T Boyd certifies that this is the approved version of the following thesis:

**Peopling of the Americas: ABO Blood Group Haplotypes as an Indicator of Native American Origins and Migration from Siberia**

_____

Chairperson: Michael H Crawford, PhD

Date Approved: July 09, 2014

## Abstract

The original peopling of the Americas has been a subject of debate among anthropologists for decades. Many molecular, archaeological and linguistic techniques have been used to assess the location of origin and number of migrations responsible for the variation observed. Two recent publications (Estrada-Mena et al. 2010 and Villanea et al. 2013) proposed a potential ancestral informative marker (AIM) in the ABO blood group region on chromosome 9 that is unique and ubiquitous in Native Americans. The marker is a subtype of the O blood group, identified as O1v542. In this study, three Beringian (Chukchi, Koryak and Aleut) and one central Siberian (Altai) population were analyzed for ABO haplotypes because of their previous connection to the original migrations into the Americas.

All four populations possessed individuals with the AIM, marking the first time the O1v542 haplotype has been observed in significant frequencies outside of Native American populations. Within and between population analyses were implemented with the four study populations and thirteen other Native American populations in order to estimate the number and timing of the migrations across the Bering land bridge and into the Americas. The findings support the Beringian Incubation Model (BIM), which proposes the ancestral population entered Beringia and paused there for a substantial amount of time because ice sheets blocked access past Alaska. During the pause, genetic drift produced unique markers, like O1v542. These markers were then transported across the Americas as founding populations moved past the receding ice sheets.

The pattern of expansion from the Beringian ancestral population, according to this study's data, supports the three-wave model. The model identifies a west-coast migratory group, an inland migratory group, and a final group that led to northern populations like the Aleut, Inuit and Eskimos. All three groups are clearly seen when looking at ABO haplotype frequencies, but a new wave back into Siberia is also insinuated from the data. This leads to the proposal of a four-wave or radial expansion model responsible for the peopling of the Americas.

# Acknowledgments

First and foremost, I must thank my incredible wife, Amanda Boyd, and our new daughter Lola June Boyd for their patience and encouragement. If it wasn't for their loving support and positive attitudes, I may never have been able to push through the times of failed experiments and poor results.

Thanks must also go out the Siberian populations represented in this study for their willing participation. Also, to Dr. Michael Crawford and Dr. Larissa Nichols for previously collecting the samples.

To the Department of Anthropology staff: Kathleen Womack, Le-Thu Erazmus and Carol Archinal, you all were a foundation of support throughout my time in the program and made my time much smoother than it would have been otherwise.

To the faculty in the Department of Anthropology, thank you for the knowledge and tools needed to advance my career and succeed wherever I go. Specifically to Dr. Jack Hofman for initially serving on my committee and providing me direction for my project. Also, to Dr. Alan Redd for the many long conversations challenging me to use my judgment and critical thinking in academic and life situations.

I am very grateful to all the members of my committee for helping me in times of stress and need. Dr. Bartholomew Dean was there to fill a committee spot on short notice and he did it with enthusiasm, which made the situation much easier. Dr. Jim Mielke helped me from day one in the program. He supplied me with research funding and projects that sustained me throughout the program, but more than that, he supplied me with reassurance and advice on how to transverse more effectively though graduate

school. Dr. Michael Crawford, my advisor and mentor, accepted me into the LBA when I had no idea what I was doing and then taught me how to persevere and succeed as both a researcher and an individual. He encouraged me to push through the troublesome times of my projects because it would make the payoff that much more fulfilling.

Thank you to my friends and family within and outside the program. To Nicholas Arnhold, Mike Guarino and Michael Davis for helping me relax outside of the lab. To Kristy Beaty and Stephen Johnson for training me and providing much needed help throughout the process. To my parents for believing I could achieve my goals. And to all my other friends and family that bolstered me; you made it all possible.

# Table of Contents

# CHAPTER 1: INTRODUCTION

Molecular genetic techniques have exploded in complexity over the past four decades. Next generation and whole genomic sequencing is becoming routine. Yet, with all the advances, two recent publications have returned to examine one of the first genetic markers globally recorded to elucidate the peopling of the Americas: the ABO blood group region (Estrada-Mena et al. 2010; Villanea et al. 2013). Karl Landsteiner discovered the ABO system in 1900 (Landsteiner 1900). Since then, it has become the most important blood group for determining successful transfusion and transplantation matches. It was also the first molecular biological marker applied to anthropological studies in humans (Roubinet 2001). Biological markers are segregating genetic traits that characterize different populations based on their presences, absence, or frequency (Crawford 1973; Crawford 2007a). Classical markers looked at proteins, white blood cells, and blood groups to understand population relationships (Crawford and Workman 1973). Modern molecular markers allow researchers to study patterns of inheritance and ancestry from the maternal (mtDNA), paternal (NRY-DNA) or both (autosomal DNA) lines. The combination of multiple markers provides robust insights into population histories.

The ABO blood group gene is an autosomal marker inherited bi-parentally. The genetic code for determining blood type is located on chromosome 9, exon 6 and 7, and was initially mapped in 1990 (Yamamoto 2000). It has since been detailed to show not just mutations altering blood type, but silent mutations that lead to different haplotypes of the well-known ABO system. The various ABO blood types are major components of

the immune system and thus allow selective advantages in different environments at different times, which make them a possible indicator of natural selection working on a population (Saitou and Yamamoto 1997). Since the different haplotypes do not incur a phenotypic change, they are selectively neutral and can be used as an indicator of mutations, genetic drift, and gene flow. The combination of all four forces of evolution that can be assessed from one marker, make them an overall good indicator of migratory history.

For this study, the ABO region will be used to assess the dispersion patterns and origins of the peopling of the Americas. The ABO system is especially valid when looking at Native Americans because they are nearly fixed for type O (Schurr 2004). The high frequency is either the result of genetic drift (founder effect) or a strong selectional barrier that early migrants crossed as they entered into the Americas. The most common haplotype of the O blood group in the Americas is O1v (Estrada-Mena et al. 2010). The O1v haplotype is seen all over the world though, so it is only a possible indicator of founder effect because of its increase in frequency. Another haplotype that is separated from O1v by a single mutation at the 542$^{nd}$ base pair is ubiquitous in Native Americans and not in any other known population. In every Native American population studied to date, this subtype has a frequency between 4% and 60% (Villanea et al. 2013). The marker has only been seen in a handful of individuals outside of the Americas, and in negligible frequencies. The distribution of this marker from the First Nations people of Canada down to the Huilliche of Southern Chile has led some to consider it as an ancestral informative marker (AIM). An AIM is a molecular marker that links many populations, suggesting a single ancestor population that split into the

current populations. Other AIMs have been studied, most notably the nine-base repeat allele (9RA) (Schroeder et al. 2007; 2009).

The 9RA, along with mitochondrial DNA (mtDNA) and non-recombining Y DNA (NRY-DNA) data, all point to a Beringia entrance into the Americas from Siberia. A land bridge across the Bering Strait would have been available for crossing from 28,000 – 11,000 before present (bp) (Lambeck et al. 2002; Meltzer 2009). Genetic connections between Native Americans and Siberians have been observed with many molecular markers. Three Siberian populations (Chukchi, Koryak, and Altai) have all of the main mtDNA haplogroups (A, B, C, D) that are found in Native Americans, and the 9RA was also found at a significant frequency in the Chukchi and Koryak populations (Perergo 2009; Schroeder et al. 2007; Tamm et al. 2007).

The mass of evidence linking Siberia to the Americas has led to the questions, not if, but when and how did the founding population(s) cross the Bering land bridge, enter the Americas, and populate throughout. Different hypotheses have been developed to address these questions. The most substantiated hypothesis is the Bering Incubation Model (BIM). This model (Tamm et al. 2007) suggests that an ancestral population moved across northeastern Siberia and then paused somewhere around the Bering land bridge. There, the population developed many unique markers through genetic drift, before continuing throughout the Americas. The population may have stopped in Beringia because much of North America was covered by two ice sheets that spanned from east to west until about 16,000 bp. From the BIM population, many have suggested a three-wave peopling of the Americas (Greenberg 1986; Reich 2012; Sicoli 2014). The first wave would have been a coastal migration across the west coast of the Americas,

since that is where the ice-sheets would have receded first. The second wave would have moved inland between the two ice sheets, and the third wave would have peopled the northern latitudes, leading to the Aleut, Inuit and Eskimo populations.

This study will examine ABO blood group haplotypes in four populations: Aleut, Chukchi, Koryak and Altai, in order to answer the following questions:

1.) Can O1v542 be substantiated as an AIM by analyzing its presence or absence?

2.) Is the high frequency of the O blood group most likely the result of natural selection or genetic drift *via* founder events?

3.) Which migration model best fits the ABO variation in Siberia and the Americas: Solutrean, One-Wave, Three-Wave, or Continuous Gene Flow?

4.) Is there a correlation between ABO haplotypes and geographic location within and outside the Americas?

If O1v542 is present, then one additional question will need to be answered:

5.) Can the presence of O1v542 be explained by more recent back migrations from Native American populations?

The four populations were chosen because of their association with the original peopling of the Americas. The Aleut are a Native Alaskan population that is thought to have been part of the last wave into the Americas after the Beringia incubation (Sicoli 2014). The Chukchi and the Koryak are located on the Siberian side of the Bering Sea, and are the two populations most consistently linked to Native Americans *via* genetic studies. The Altai are a native central Siberian population. They are present in the study since they have all the mtDNA and NRY-DNA markers found in the Americas. They will also provide a reference to see how far back the O1v542 haplotype may have evolved. In

order to strengthen the results, the four populations from this study will be combined with 13 previously analyzed populations from North America, Mesoamerica and South America. The combined analyses will create a comprehensive picture of ABO haplotype diversity in Siberia and the Americas.

## CHAPTER 2: BACKGROUND

Discussion over the original peopling of the Americas has peaked the interest and permeated research in many of the subfields of anthropology. Archaeologists, linguists, osteologists, and genetic anthropologists critically inspect both past and present Siberian and Native American populations, languages and artifacts in an attempt to establish a conclusive dispersal origin and migratory route throughout the Americas. Much of the evidence coalesces with the Siberian origins model dated between 13,000 and 16,000 before present (bp); but establishing a consensus view and pinpointing the precise location and path(s) of migration(s) has proven difficult. The primary hypotheses attempt to distinguish between a post-Last Glacial Maximum (LGM) ice-free corridor (IFC) migration leading to the Clovis culture; a pacific-coast migration, one or two millennium earlier leaving sites all the way to the coast of South America; and a potential pre-LGM migration, establishing the pre-Clovis sites across the eastern United States. Other theories such as a French Solutrean origin of Native Americans have been posited but are not seriously considered by the majority of researchers (Bradley 2004). In fact, tracing migrations simply through lithic technology and material culture, which is the basis of the Solutrean hypothesis, is no more exact today than it was 100 years ago (Waguespack 2007:66).

The trail agreed upon by the linguistic, craniofacial, genetic and archaeological data proposes a south-central Siberian origin in the Altai/Baikal region that then progressed through western Beringia, into eastern Beringia, and either split into two migrations—one coastal and one inland—or remained a single pulse into the Americas.

The linguistic and genetic evidence are important in establishing general relationships and potential areas of origin, but the dating methods (glottochronology and mutation rates) do not always provide narrow ranges or conclusive dates necessary for predicting migratory dispersions. Likewise, one cannot assume that migrating groups left behind genetic breadcrumbs for modern people to trace their paths. It has been at least 12-15,000 years since the original migrations would have occurred, so there is no assuredness on whether a current group looks like an original founding group, a later migrating group, or a mixture of multiple groups. For instance, if a coastal and an inland group came in contact and had significant gene flow, then the original differences that could trace a two-wave migration would have been lost. Because of such potential confounding factors, identifying a mixture of modern and ancestral genetic and linguistic data along with an assortment of archaeological sites and radiocarbon dates builds a more holistic and informative picture.

**2.1 Climate**

The principal regulatory force that controlled the timing and paths of expansion in the first peopling of the New World was the Last Glacial Maximum. Occurring between 24,000-20,000 bp, it resulted in two massive ice sheets, the Laurentide and Cordilleran, which converged by about 28,000 bp and did not provide an accessible path to the Americas until 16,000 bp. But, during the LGM, roughly 5.2% of the world's water was frozen on land (about half in the Laurentide ice sheet alone), and that frozen mass drew down the seas 120 ± 10 meters below their present level (Lambeck et al. 2002). For the Bering and Chukchi seas to become dry land and allow access between Siberia and the

Americas, the seas only had to fall about 50m, exposing the continental shelf. During LGM, the Bering Land Bridge, roughly 1,000km wide became traversable (Meltzer 2009:34). It remained passable from about 28,000 to 11,000 bp, but as mentioned earlier, the North American ice sheets were impeding southward movement. Then, between 16,000 and 15,000 bp, the de-glaciation process slowly opened the pacific coast for what could have been the first movement beyond Beringia and into the Americas. About 2,000 years later, between 14,000 and 13,000 bp, the area between the two ice sheets began to retreat and open a new entryway into the Americas.

The non-glacial climate and landscape during the LGM along the Bering Land Bridge also played a major role because of the range of animal species sustained. The environment during the LGM was significantly drier and colder than modern times, and the extensive animal remains reveal that the same species roamed between the land bridge on both the Siberian and American sides. Large grazing animals and their predators populated the area to the extent that some scholars speculate it resembled a colder and drier African savanna (Guthrie 1990). More recent analysis of pollen records reveals LGM Beringia to be in-between the current tundra and a savanna-like steppe in composition and diversity (Anderson et al. 2004). Either way, the range of megafauna exposed it to be a primary location for hunter-gatherer bands to locate a large quantity of calories and potentially those hunter-gatherers then followed the herds across the Bering Land Bridge and into the New World. The "piggybacking on megafaunal movements/extinctions" hypothesis has taken much criticism, however, because faunal remains at many sites point to a wide diversity of dietary sources, both plant and animal. Other critiques of this hypothesis stress the increased difficulty for hunting in new

landscapes, so why would large game hunters move extreme distances to new environments where prey and landscape changes complicate previously learned strategies? This would be especially true when groups arrived south of the ice sheets where food was plentiful in both animal and plant form, but other traditional motives for migration would not be applicable in this situation: over-population, exile, warfare, or trade (Meltzer 2009: 215). Harsh climate changes and chasing calories are the best possible reasons for movement; so reviewing the fossil record across Siberia and Beringia and how the dates correspond to the LGM makes a great strategy for deciphering hunter-gatherer migration patterns.

### 2.2 Archaeological Sites

*See Figure 1 for a map of all sites discussed*

There are few accepted pre-LGM and LGM sites in Siberia and Beringia to point to an origin location, time, or path, and the few best-excavated sites leave a major gap of 18,000 years between them. There are other sites that will be discussed later, which can fill parts of this gap as they are further excavated and substantiated. The oldest, most pertinent site in Siberia is the Yana Rhinoceros Horn Site in western Beringia dated to 32,000 cal. Bp (Goebel et al. 2008). This site is the first to demonstrate that people adapted to handle life in the far north shortly after arriving in Siberia (Hoffecker and Elias 2007). The site contains a well-preserved frozen layer of fauna and cultural artifacts. The most interesting aspect of the site is the Clovis-like carved wooly rhino foreshafts and bifacial chipped stone tools. These remains point to an American ancestral site, but lack of other sites chronologically or geographically moving toward the Americas makes this site currently not as impactful. The site does demand the

question of whether the group retreated back to South-central Siberia during the LGM or if they moved east toward Alaska, but without further discovery we cannot establish an answer. The closest dated site in Beringia is the Swan Point Site in Central Alaska/Eastern Beringia dated to 14,200 cal. bp. The site contains microblades and stone implements that resemble three central Siberia sites: Nizhinii Idzir (20,000 bp), Kaergas Cave (18,000 bp) and Diuktai Cave (16,000 bp) (Bever 2006). The confluence of these three sites with the Swan Point site suggests a possible migration path, but there are no sites in western Beringia, leaving a spatial gap this time instead of a chronological gap. If these sites do represent the path leading to the peopling of the New World, then it suggests that the people of Beringia retreated back to southern Siberia during the LGM because they were not equipped to deal with such dry cold weather. It also would indicate that there was no pre-LGM movement into the Americas, thus further denying the legitimacy of the American sites dated past 15,000 bp.

After the Swan Point site, the record becomes more complicated. The Nenana complex of central Alaska—close to Swan Point—is dated between 13,800 bp and 13,000 bp, but it lacks the microblades and burins of its contemporary and instead only has bifaces and unifaces made on blades and flakes (Goebel et al. 2008). The other site dating to this time period is the Ushki site in the Kamchatka peninsula of Siberia, and its complex looks very similar to the Nenana complex, potentially suggesting another migration from Siberia to Alaska. A northwest Alaskan site, Sluiceway-Tuluaq, is dated similarly to Nenana and Ushki but possess technologically distinct artifacts such as lanceolate bifaces (Hoffecker and Elias 2007). It is not until after 13,000 bp that the microblade and burin technologies reappear in Alaskan complexes; but even these

cannot be directly traced to Clovis or pre-Clovis tools, as is expected with the difficulty of using material culture to determine migratory routes.

As mentioned earlier, there are sites that fill some of the previous gaps in Beringia but are not universally accepted. They are mentioned here because of their relevance to the discussion, but with the understanding that more work needs to be undertaken to validate their status. Bluefish Cave is the most convincing early site located in the northern Yukon. The site is associated with microblades, burins, and flakes and is dated to about 16,000 bp (Morlan 2003). The site suggests an earlier occupation of eastern Beringia than previously found. Three other sites have been discovered in western Beringia: Berelekh (16,100 bp) near the Chukchi Peninsula, Ikhine 2 (24,300 bp) and Verkhine-Troitskaya (22,000 bp), both in the Yukutia along the Aldan River (Pitblado 2011). If the site dates hold up, they will disprove the theory that Beringian populations retreated to southern Siberia. This would reinforce the major hypothesis that the people who populated the Americas stalled in Beringia throughout the LGM, where they developed some of the unique genetic markers found across all Native Americans (discussed further below). In Siberian and Beringian assemblages we do see the same broad classes of tools and technologies that are present in early America, whether Clovis or pre-Clovis—bifaces, points, end and side scrapers, gravers, blades, and flake tools—so the relationship is apparent, but the connection is not fully known (Meltzer 2009:189). The earliest sites in the Americas reveal the continued migration path(s) after leaving the Beringia/Alaska region.

The Pacific coast ice sheet was the first to recede, opening the first possible route to the continental United States, Central America, and South America. The site, which

started the coastal migration hypothesis and remains one of the most thoroughly excavated sites in the New World, is the Monte Verde site in Chile. The site is dated to 14,600 cal. bp, 400 years before Swan Point and over a millennium before the interior ice-free corridor opened. As Meltzer (2009) explains:

> As for Monte Verde (which is credible), its toolkit seems very different from Clovis in North America, and from Clovis-age South American archaeological complexes (again, there is no Clovis, strictly speaking, in South America). Perhaps Monte Verde and Clovis represent two distinct archaeological traditions and separate migratory pulses, as opposed to an outgrowth of the same historically related colonization. At the moment the call could go either way (201).

Monte Verde demonstrates year-round use of marine resources through the finding of nets, nine marine algae species, and seaweed representing different seasons and coastal settings (Dillehay et al. 2008). The intimate knowledge of the sea and its resources suggests a coastal migration and strongly implies an association of boat-travel to account for the speed necessary to reach the site so quickly after the coastal passage opened. It is understood that although Monte Verde represents an early presence, it is assuredly not the earliest. There are sites such as Taima Taima north of Monte Verde that has a very similar assemblage and is dated slightly earlier, but questions still surround its dating accuracy. Monte Verde also has another section of the site dated to about 30,000 bp, but it has not been substantiated. The next oldest coastal site is a recent discovery of coprolites at Paisley 5 Mile Point Cave, just inland in Oregon. The coprolites were genetically analyzed and proved to be human with similar haplogroups as other Native Americans. The coprolites are dated to 14,300 cal. bp. The cave is located on the remains of a Pleistocene lake that would have been accessible from the Pacific Ocean, again suggesting a coastal push. The other valuable coastal sites are Arlington Springs and Indian Sands, dated to 13,100 and 12,500 cal. bp, respectively. Arlington

Springs provided human skeletal remains, and even during the Pleistocene it would have only been accessible by boat. Although there has never been evidence of sea-craft dating before 10,000 bp in the Americas or Asia, many of the coastal sites along with the speed necessary to populate the sites indicate boat use. Indian Sands is another site along the Oregon coast with burned and unburned mussel shells (Pitblado 2011). One explanation suggested for the rapid coastal spread is the process of scouting or hunters pursuing terrestrial game while probing new territories (Meltzer 2009: 223). If this is the case then some of the oldest archaeological sites may not represent settlements but temporary encampments of scouting parties. A site such as Monte Verde, however, where a year-round occupation has been established, does not appear to be a small scouting party encampment but at least a multi-year settlement.

An inland migration is nearly ubiquitous in the archaeology world. This is because of the expanse of Clovis sites and the multiple pre-Clovis sites throughout central and eastern United States. The only other explanation is that the coastal migration mentioned above spawned all the inland sites across the Americas. The Clovis (dated 13,200 to 12,800 cal. bp) and pre-Clovis cultures consist of a completely different subsistence strategy—terrestrial hunters and gatherers—that would have taken much time to develop, especially in unfamiliar landscapes, and as Meltzer mentioned above, the technologies are vastly different. Clovis and pre-Clovis sites boast lanceolate fluted projectile points, end scrapers, bifaces and unifaces, and ivory/bone/antler foreshafts (Goebel et al. 2008). It is difficult to establish an order of these sites because of either their closeness in time or their vast diversity in dates, but looking at the location of sites in relation to the ice-free corridor is evidence enough that

the passage between glaciers was used (figure 2 and 3). Many of the pre-Clovis sites and dates are still being discussed and not always civilly. Page-Ladson (14,400 bp), Schafer and Hebior (14,200 bp), and Meadowcroft (15,200 to 13,400 bp) are all well excavated sites that receive much criticism because they complicate the entrance time(s) and path(s). All of these sites can still be reasoned into either a part of the coastal expansion as the people followed rivers and streams throughout the United States to the east coast, or as a very early and rapid part of an inland expansion. Other sites such as Cactus Hill (20,000 to 16,000 bp), Topper (20,000 to 16,000 bp), La Sena and Lovewell (22,000 to 19,000 bp) are not as irrefutable. Dating of Cactus Hill and Topper both relied upon charcoal samples recovered as isolated fragments from other areas of the dig site and indicate translocation. The fact that there are assemblages in stratigraphic layers with charcoal dated to this older time period is still compelling, but until more conclusive dates, these sites do not alter the pre-Clovis peopling hypothesis. La Sena and Lovewell are less compelling because there are no stone tools or butchering artifacts associated with the site, only faunal remains that possibly suggest intentional fracturing and flaking. So without undeniable material culture, these two sites cannot at this time play a part in determining a chronology. There is still no site accepted that dates to a pre-LGM population, although there are more and more sites challenging established dates of the initial peopling. As more sites point to a pre-LGM migration, a new conundrum of how people got to the Americas when there was no direct path will need to be answered. The only possible solution to still support genetic evidence would be a major sea voyage from somewhere in Siberia, but there is even less evidence for this than a pre-LGM

peopling. Further analysis utilizing linguistics and genetics are necessary to piece together the archaeological puzzle of the peopling of the Americas.

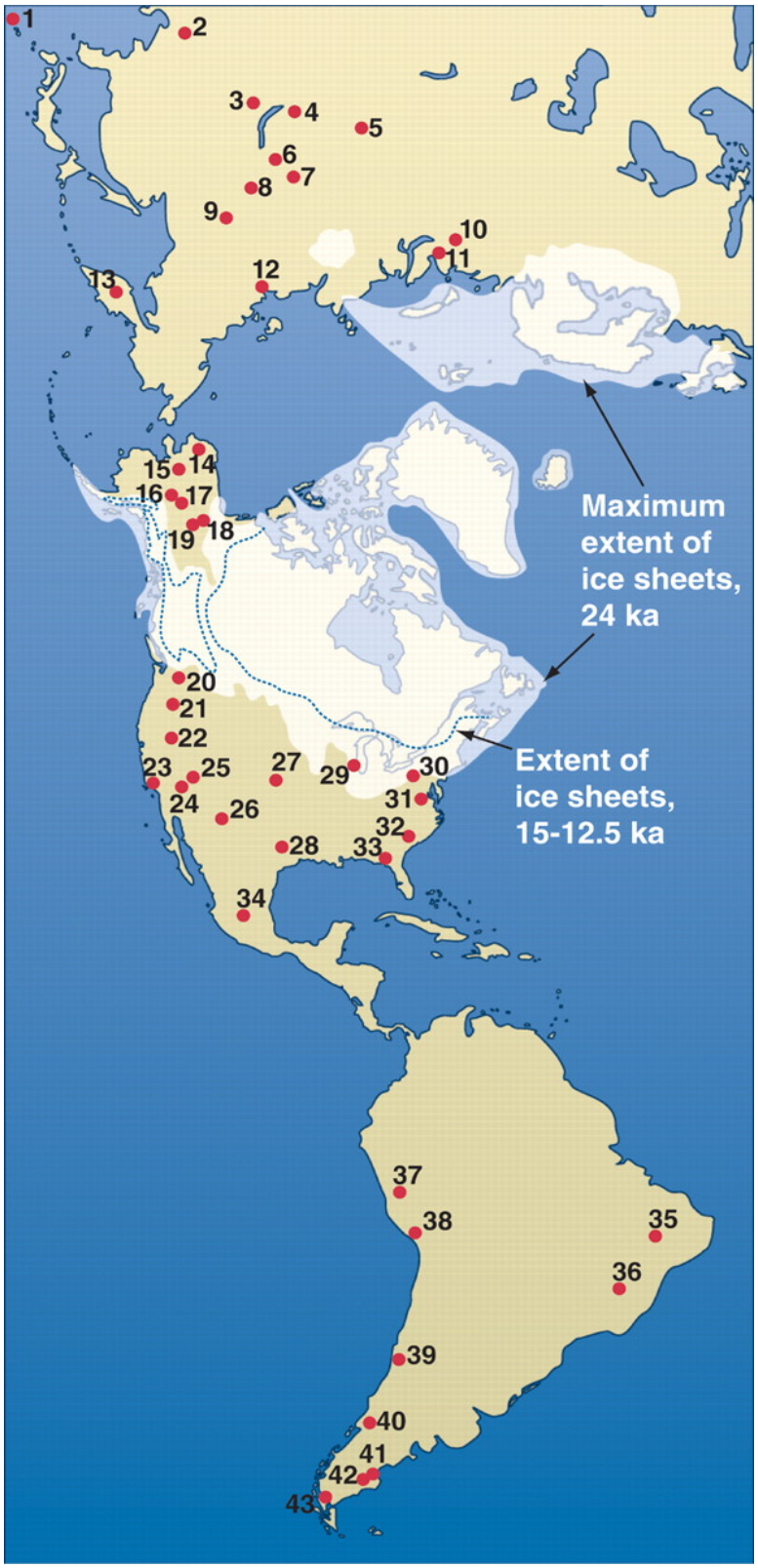



**Figure 1:** Map showing location of archaeological sites mentioned in text (1, Yamashita-cho; 2, Tianyuan Cave; 3, Studenoe-2; 4, Mal'ta; 5, Nizhnii Idzhir; 6, Alekseevsk; 7, Nepa-1; 8, Khaergas Cave; 9, Diuktai Cave; 10, Byzovaia; 11, Mamontovaya Kurya; 12, Yana RHS; 13, Ushki; 14, Tuluaq; 15, Nogahabara I; 16, Nenana; 17, Swan Point; 18, Old Crow; 19, Bluefish Caves; 20, Kennewick; 21, Paisley Caves; 22, Spirit Cave; 23, Arlington Springs; 24, Calico; 25, Tule Spring; 26, Pendejo Cave; 27, La Sena and Lovewell; 28, Gault; 29, Schaefer, Hebior, and Mud Lake; 30, Meadowcroft Rockshelter; 31, Cactus Hill; 32, Topper; 33, Page-Ladson; 34, Tlapacoya; 35, Pedra Furada; 36, Lagoa Santa; 37, Pikimachay; 38, Quebrada Jaguay; 39, Quebrada Santa Julia; 40, Monte Verde; 41, Piedra Museo; 42, Cerro Tres Tatas and Cuevo Casa del Minero; 43, Fell's Cave). (Goebel et al. 2008)

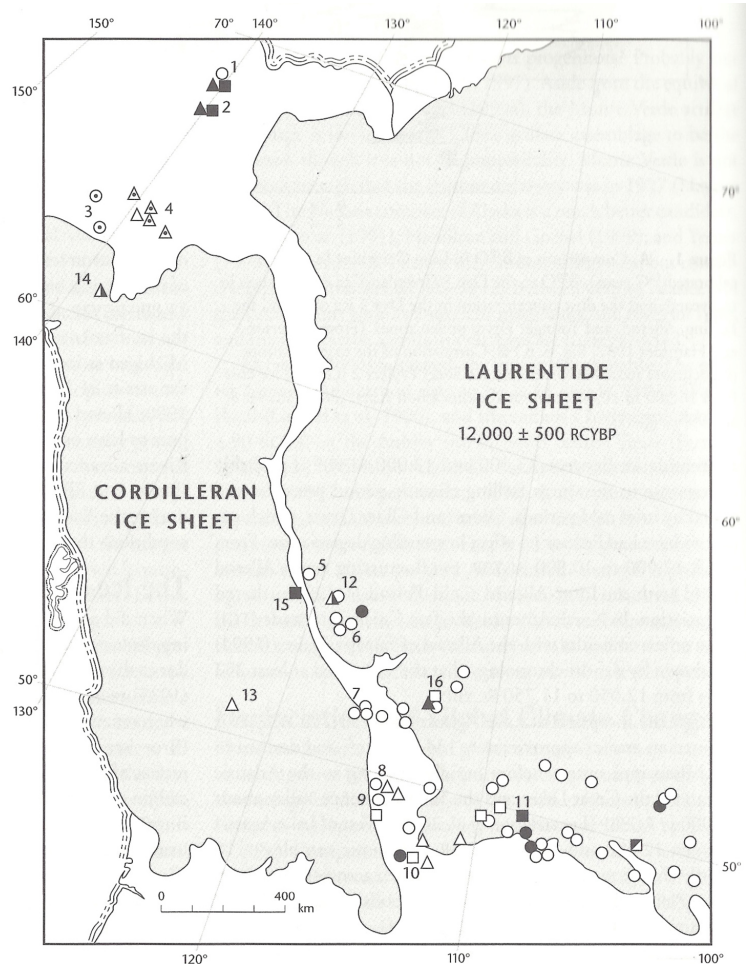



**Figure 2.** Generalized map of ice-free corridor at 13,000 RCYBP after Dyke and Prest (1987).

- ● Clovis point
- ○ other fluted points
- ⊙ Nenana sites

radiocarbon-dated bones:
- ▲ large mammal >11,500 RCYBP
- △ large mammal 11,500–11,000 RCYBP
- ◬ arctic ground squirrel bone or nest >11,000 RCYBP

radiocarbon-dated wood or organic matter:
- ◭ minimum age for deglaciation
- ■ mammoth bone ≥11,500 RCYBP
- □ mammoth bone < 11,500 RCYBP
- ◩ Clovis mammoth-bone point or rod

1 Old Crow basin
2 Bluefish Caves
3 Tanana River Nenana sites
4 Dawson-Klondike area
5 Charlie Lake Cave
6 Birch Hills
7 fluted-point concentration in upper Pembina Valley
8 Vermillion Lakes site
9 Sibbald Flat
10 St. Mary Lake mammoth tracks
11 Kyle mammoth
12 Boone Lake
13 Lochnore Creek, Fraser River Valley
14 Kluane Lake, Yukon Territory
15 Portage Mountain Dam mammoth tusk
16 Clover Bar gravel pit, east Edmonton, Alberta

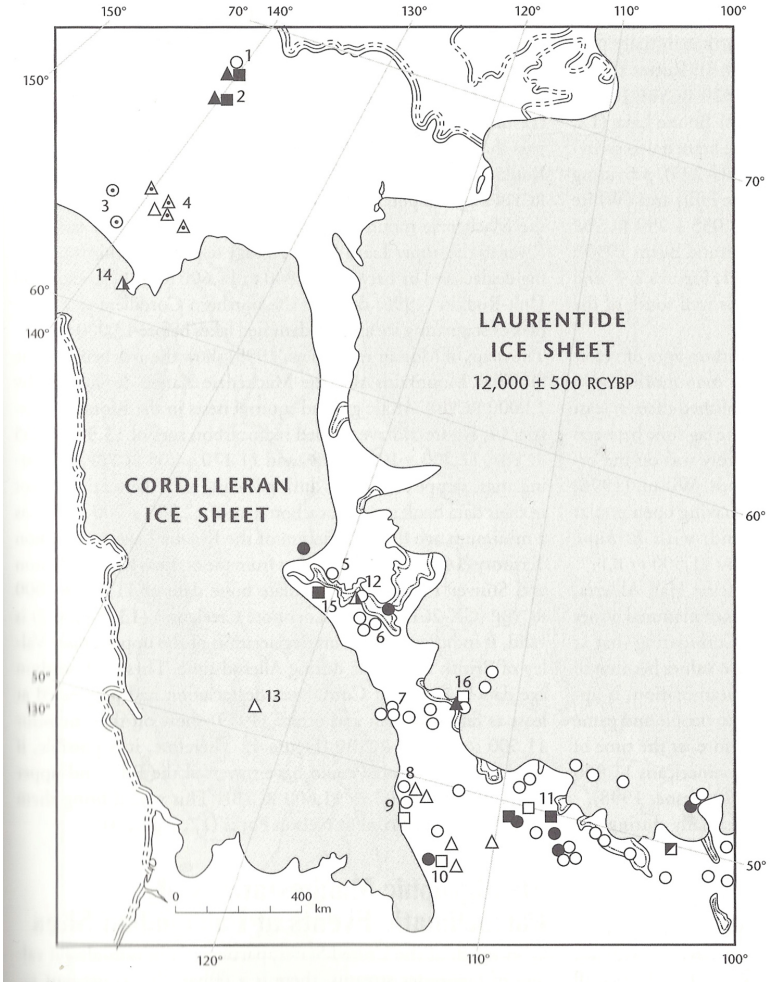



**Figure 3.** Generalized map of the ice-free corridor at 12,000 RCYBP, modified after Morlan et al. (1999) and Jackson and Duk-Rodkin (1996).

- ● Clovis point
- ○ other fluted points
- ⊙ Nenana sites

radiocarbon-dated bones:
- ▲ large mammal >11,500 RCYBP
- △ large mammal 11,500–11,000 RCYBP
- ◬ arctic ground squirrel bone or nest >11,000 RCYBP

radiocarbon-dated wood or organic matter:
- ◭ minimum age for deglaciation
- ■ mammoth bone ≥11,500 RCYBP
- □ mammoth bone < 11,500 RCYBP
- ◩ Clovis mammoth-bone point or rod

1 Old Crow basin
2 Bluefish Caves
3 Tanana River Nenana sites
4 Dawson-Klondike area
5 Charlie Lake Cave
6 Birch Hills
7 fluted-point concentration in upper Pembina Valley
8 Vermillion Lakes site
9 Sibbald Flat
10 St. Mary Lake mammoth tracks
11 Kyle mammoth
12 Boone Lake
13 Lochnore Creek, Fraser River Valley
14 Kluane Lake, Yukon Territory
15 Portage Mountain Dam mammoth tusk
16 Clover Bar gravel pit, east Edmonton, Alberta

## 2.3 Linguistics

The linguistic evidence for establishing migratory routes and timelines for the peopling of the Americas has remained essentially unchanged since Greenberg et al. (1986) proposed a three-wave model: the earliest wave procuring the greatest distribution across North, Meso and South American groups, the second wave forming the Na-Dene speaking groups between Alaska and the Pacific Northwest, and the third producing the Eskimo/Aleut groups. Most of the conclusions were speculative because the traditional comparative method of historical linguists has a limit of 8-10,000 years before the connections become too jumbled to decipher (Greenhill et al. 2010; Nichols 1992). However, new computational phylogenetic tools (Campbell 2011; Diamond 2011) have recently been developed and employed in linguistics to rigorously test the three-wave hypothesis and establish connections between different language groups.

One study established a connection between the Yeniseian language group of Siberia and the Na-Dene speakers and suggested the Native American origin occurred in Central or Western Asia using multilateral comparison of lexical items (Ruhlen 1998). A stronger case for the connection between the Siberian and Na-Dene groups was further established by new linguistic reconstruction methods (Vajda 2010). Sicoli and Holton (2014) used Bayesian methodology on the same data set to support an Out-of-Beringia dispersal and not an Out-of-Central/Western-Asia dispersal as conjectured by Ruhlen (1998). According to the Bayesian data, the original populations had to pause in the Beringia area for long enough to allow the language to become substantially distinct from languages spoken in other parts of Asia. This "Beringian Standstill" was first claimed using mtDNA data, as will be discussed below (Tamm et al. 2007). The study by

Sicoli and Holton (2014) demonstrates that the Yeniseian groups were not the origin population, but a result of a back migration from the Beringian population. The new study adds to the complexity of the peopling by insisting that the migration occurred in both directions from the founding location somewhere in Beringia. With the implementation of new linguistically statistics, many more studies are bound to help clarify the linguistic side of the original migrations.

## 2.4 Molecular Markers

### *2.4.1 ABO Blood Group*

In order to narrow the origin regions and dates of group splits, genetic data provides a more objective depiction of the relationships between current groups from multiple regions and with archaic groups if ancient DNA (aDNA) is available. One major form of analysis that has been used to examine population affinities since the early 1900s is ABO blood grouping. Recently, ABO typing evolved to look at genetic subtypes — or haplotypes — to further identify the different mutational patterns leading to the different types and which populations are more closely related in those mutations. Mourant et al. (1976) composed a cumulative book of blood group distributions that revealed interesting global clines. While type O was observed as the most prominent globally (~63%), as populations moved east and north through Siberia and then into the Americas, the percentage of O phenotypes increased from about 62% in Siberia to >85% in North America and finally >95% in South America. Some initial speculation about the reason for Native American populations being nearly fixed for type O pointed to the intensive selection the majority of natives went through after European contact.

Recently, multiple ancient DNA studies have looked at blood groups and determined that pre-Columbian North Americans Type O frequency was about 96.7% (Halverson and Bolnick 2008), while pre-Columbian South American Type O frequency was 100% for all ancient individuals studied (Georges et al. 2012). This pattern is not found in any other populations (Swerdlow et al. 1994, Molnar 2002, Llop et al. 2006). The A and B alleles are still found at significant rates in the Siberian, Na-Dene and Alaskan Aleut and Eskimo populations (Szatharmy 1979). The O allele evolved from the A allele by a deletion at position 261 on exon 6 of the ABO gene. The deletion causes a premature stop codon leading to a protein that lacks glycosyltransferase function (Gagneux and Varki 1999). Since the mutation causes a non-functional protein, other mutations can occur in the gene without impacting the translation (Yip 2002). There are many haplotypes of the O allele that can be seen across the globe. Table 1 reveals various mutations on exon 6 and exon 7 leading to different subtypes of the ABO alleles.

A recent study by Estrada-Mena et al. (2010) of modern South American natives revealed the potential for an ancestral ABO haplotype that could be used to trace initial migrations into the Americas. The haplotype is O1v542 (O1vG542A). O1v is the most prevalent haplotype in genetic ABO studies in the Americas: about 75% of type O individuals (Georges et al. 2012). Not all of these studies, particularly the ancient DNA studies, looked for the 542 mutation, but in thirteen different Mesoamerican and South American populations O1v542 reached between 4 and 60%, while none of the Asian populations of Japan, China or Korea had the haplotype (Olsson et al. 1998, Roubinet et al. 2001, Barjas-Castro et al. 2003, Llop et al 2006, Estrada-Mena et al 2010, Villanea et al. 2013). One individual in a European sample and one individual in a Middle-Eastern

*Table 1. Nucelotide sequence variation in exons 6 and 7 of ABO haplotypes*

| | Exon 6 | | Exon 7 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Allele** | **261** | **297** | **467** | **498** | **526** | **538** | **542** | **646** | **657** | **681** | **703** | **771** | **796** | **803** | **829** | **930** |
| A101 | G | A | C | C | C | C | G | T | C | G | G | C | C | G | G | G |
| A102 | * | * | T | * | * | * | * | * | * | * | * | * | * | * | * | * |
| B101 | * | G | * | * | G | * | * | * | T | * | A | * | A | C | * | A |
| O1 | - | * | * | * | * | * | * | * | * | * | * | * | * | * | * | * |
| O05 | - | G | * | * | * | * | * | * | * | * | * | * | * | * | * | * |
| O1v | - | G | * | * | * | * | * | A | * | A | * | T | * | * | A | * |
| O1v542 | - | G | * | * | * | * | A | A | * | A | * | T | * | * | A | * |
| Ov7 | - | G | * | * | * | * | * | A | * | * | * | T | * | * | A | * |
| O32 | - | G | * | * | * | T | * | A | * | A | * | T | * | * | A | * |
| O33 | - | G | * | T | * | * | * | A | * | A | * | T | * | * | A | * |

population demonstrated this marker, but the extremely low frequency in both populations could have been a result of recent gene flow from Native Americans to those areas (Yip 2000, Yip et al. 2006). As an ancestral marker, this ABO haplotype could reveal a specific founding population for the Americas. If analysis of native Siberian populations shows a distinct group with a similar frequency of O1v542, while other Siberian populations do not possess the haplotype, it will be robust evidence for the initial population and route into the Americas. Even the current ABO data reveals a much stronger connection with Siberia and East Asia in the frequency of O1 and O1v haplotypes than with Europeans who have much greater percentages of A, B, and AB blood types and O2 and O3 haplotypes. Beyond only looking at ABO blood group haplotypes, many other genetic markers provide substantial evidence for the BIM and multi-wave peopling of the Americas that is supported and furthered by many archaeological sites.

*2.4.2 mtDNA, NRY, and Autosomal*

Genetic evidence points to northeast Asia as the immediate point of origin into the Americas. Prior to that launching point, the earliest inhabitants hailed from more southern regions of Asia, but even before that they descended from a population dispersing from Africa by 100+ thousand years ago (kya) (in the Arabian Peninsular) and appearing in Asia by 40 kya (O'Rourke 2008). The evidence surrounding this prior region of origin before reaching Beringia is more controversial in regards to the degree of specificity scientists are willing to accept. In fact, the hypothetical locations of interest are more jumping-off-points or areas of stagnation that were then followed by pulses

toward the Americas than they are "origin" locations. The long list of "source" regions discussed by geneticists varies from south and south-central Siberia (specifically the Altai mountains and Lake Baikal region) as the most prominent hypotheses (Bortolini 2003, Derenko et al. 2007, Dulik et al. 2012, O'Rourke 2009, Schurr 2004) to eastern Siberia (Wang et al., 2007), opponents of eastern Siberia (Rubicz 2010), Asia in general (Fagundes et al. 2008), as well as Mongolia (Malhi and Smith 2002) and many others. Although this appears to be an exceptionally broad range of possibilities, considering the fact that the current Far Eastern Area District (FEAD) of Russia, which contains the Beringia area, has one of the lowest population densities at 1.1 peoples/km$^2$, and would have been exponentially lower with hunter gatherer bands in the Paleolithic, the potential source pool that all these origin variants are drawing from is minute (Pitblado 2011). The primary genetic markers used to assess the dispersal points and the number and routes of migrations reveal some consensus across the board.

MtDNA is the most studied evidence for population relationships between Siberia and the Americas. The number of Native American haplogroups has slowly increased over the past two decades as more samples and more efficient methods have been used. The four pan-American haplogroups of A2, B2, C1 and D1 are seen across the North and South American continents, but five minor lineages have more recently been deciphered: C4c, D2a, D3, D4h3, and X2a (Perergo 2009). The major four haplogroups are dispersed evenly across the Americas, which is one indication of their uniform

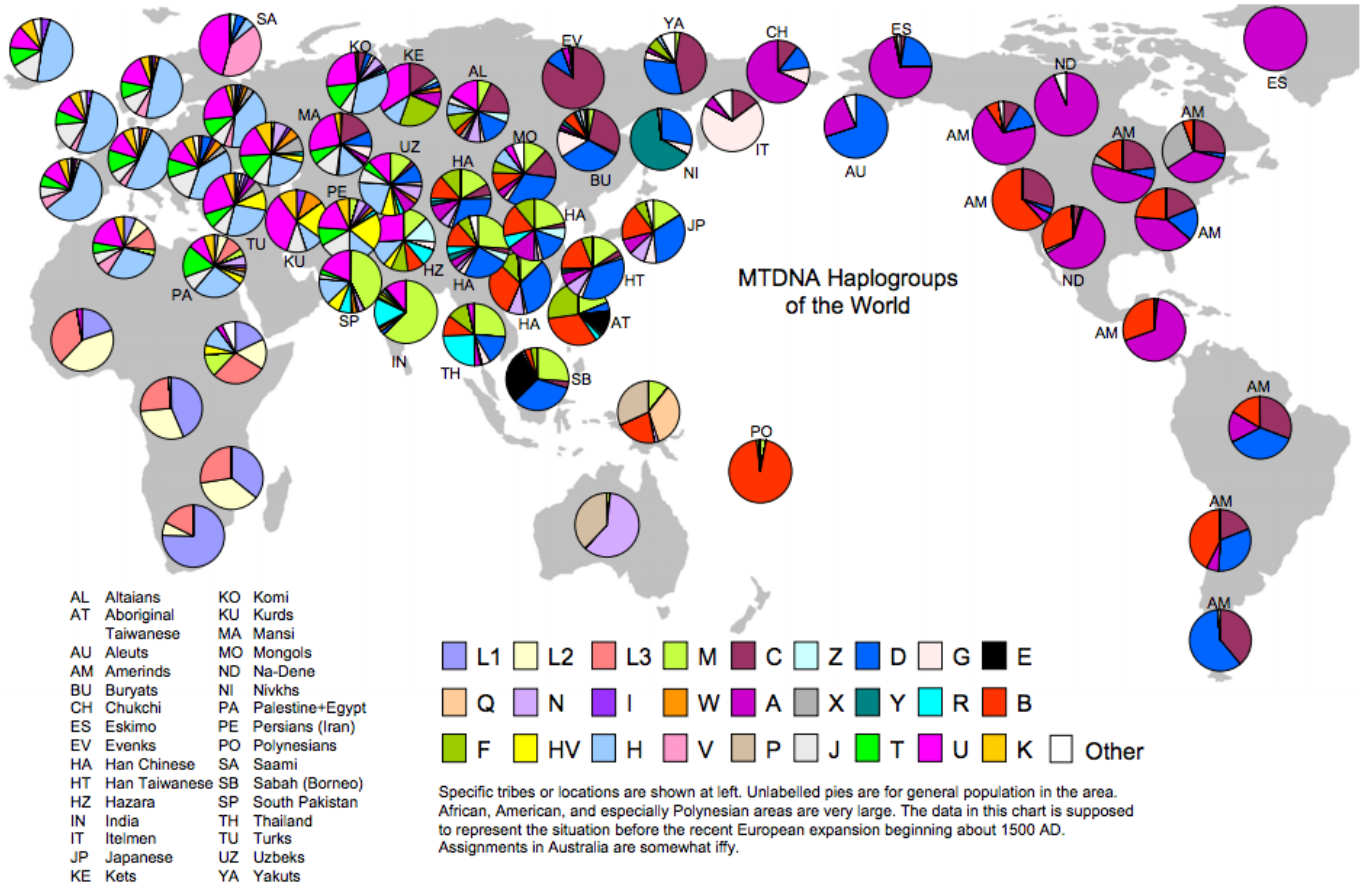



**Figure 4:** Worldwide mtDNA distributions. Figure taken from McDonald JD 2005. http://www.scs.illinois.edu/~mcdonald/WorldHaplogroupsMaps.pdf

ancestry, and they are also all found in certain Siberian populations such as the Altai (Figure 4). Variants of A and X are seen in Europe, but they are different mutational forms, so Eastern Asia is the only place that could possibly spawn the specific mtDNA diversity of Native Americans. More refined analysis by Tamm (2007) looked at three subclades of haplogroup C seen in Native Americans—C1b, C1c and C1d— for Beringian/Siberian distribution. All three subclades were found to be absent in Asians, with coalescence times of 13,900 ± 2,700 bp, revealing sufficient isolation occurred to allow these subclades to split from their Asian sister-clades. This Beringian isolation model suggests that pre-LGM groups became isolated in Beringia from about 30,000 bp until deglaciation began around 15,000 bp. During this 10,000 – 15,000 year isolation,

due to glaciers inpeding their path, the population that entered the Americas developed distinct variants from the other Siberian populations.

The nine-base repeat allele DYS199t (9RA) proposed by Schroeder et al. (2007, 2009) is another marker of this hypothesis. This distinct autosomal marker is found only in Native American populations and at a rate greater than 30% across all groups. Two Beringian groups, the Chukchi and Koryaks, also share this mutation at a similar rate to that of the Native American groups. It is thought that the 9RA along with the mtDNA marker 16111T that distinguishes lineage A2, and the SNP that distinguishes Y-chromosome lineage Q-M3 are evidence of an ancient gene pool that included the ancestors of modern inhabitant of Western Beringia and America (Underhill et al. 1996). The presence of all these markers in the three language groups discussed earlier (Amerind, Na-Dene and Aleut/Eskimo) points toward the Beringia Incubation Model where these traits developed before the expansions that peopled the Americas.

Crawford (2007b) compared Aleutian NRY, mtDNA, and autosomal STRs to eastern Beringian populations, and concluded the mtDNA haplogroup lineages and STRs reveal an ancient barrier (isolation-by-distance) between Bering Island Aleuts and Kamchatkan populations. Rubicz et al. (2011) further analyzed these data and concluded that the Kamchatkan populations are not remnants of an origin population as suggested by the Beringian Incubation Model proposed by Tamm (2007). The authors contend that the similarities between the Beringian populations and the Aleutians are owed to back migrations. This conclusion does not hold for other markers such as the 9RA. Back migration of the 9RA allele is not robust enough to prove the high frequency relationship between the two Beringian populations and all Native Americans. The extent of back-

migration necessary to create a 30% frequency would have established sufficient similarity that the data collected by Rubicz et al. (2010) would have revealed a stronger relationship. A more likely explanation of the variation that is seen between the groups is isolation plus genetic drift and a separate back migration as mentioned earlier. Multiple studies that used the Bayesian Skyline Plot (BSP) method on mtDNA datasets have shown a period of bottleneck followed by a large population expansion after the LGM between 19,000 and 16,000 bp (Elias 2001; Fagundes et al. 2008). The Beringia bottleneck would create a narrow spread of mtDNA, NRY, blood groups, and autosomal markers such as the 9RA that are ubiquitous in Native Americans. The genetic divergence between Bering populations and Native Americans is expected because of the extreme climatic selection and the time separation of migratory waves. Most of the data from Rubicz et al. (2010) reveals as much a separation between the Alaskan Aleuts and other Native American groups as it does between the Aleuts and the Bering groups, which fits the narrative. The founding population appears to have split into two groups: one group remaining in central/eastern Beringia that is potentially ancestral to all Western Beringians, such as the Chukchi, and one group that then pushed further into Alaska in response to the cessation of the LGM. If the BSP method is accurate, the latter group would have experienced massive expansion and split further into the groups that led to the three-wave model. The pattern of expansion followed by stasis would have created more diversity within haplogroups that are not seen in Asia but are seen either in the Alaska region or across the Americas. The extreme population expansion is also one of the primary candidates for why such a rapid movement into the Americas and all the way into South America would have occurred after the coastal path opened. The

present day haplogroups C and D support the three-wave hypothesis because dating suggests the initial clades expanded immediately before the LGM in Eastern Asia, while many of the northern Asian variants began expansion towards the end of the LGM, and the American C and D haplogroups expanded about the same time just following the LGM, demonstrating multiple splits and isolations within haplogroups (Derenko et al. 2010).

The three-wave hypothesis does not ignore that more recent admixture from Russian sources may have occurred and could have skewed the picture, but recent admixture has affected the NRY data much more significantly than other data sources. As Rubicz (2010) shows, there are no longer clear distinctions between Kamchatkans, Bering Islanders, and Aleuts in their NRY. The Y-chromosome data from the Americas is similar to the mtDNA in that there are few haplogroups present: C, P, Q, and R. Again, Southern Altaian populations are the only populations that possess all four haplogroups (Dulik 2012) (Figure 5). A low-level NRY-SNP reveals a similarity between southern Altaian and Native American Q and C haplogroups. Recent in-depth analysis of haplogroup Q for new mutations and resolution has made it the best source of phylogenetic analysis of the Native American Y-haplogroups. The high-resolution allows for more accurate TMRCA dating. For instance, recent southern Altaian Q-M346 also possesses L54, a SNP shared by Q-M346 Native Americans. This reveals the probable ancestral haplogroup from which Q-M3 later developed and is dated to around 22,000 bp based on one interpretation of the NRY mutation rate (Dulik 2012). Principal Components Analysis between multiple Asian and Native North American groups revealed a distinct cluster between the Native North Americans (Chipewayan and the

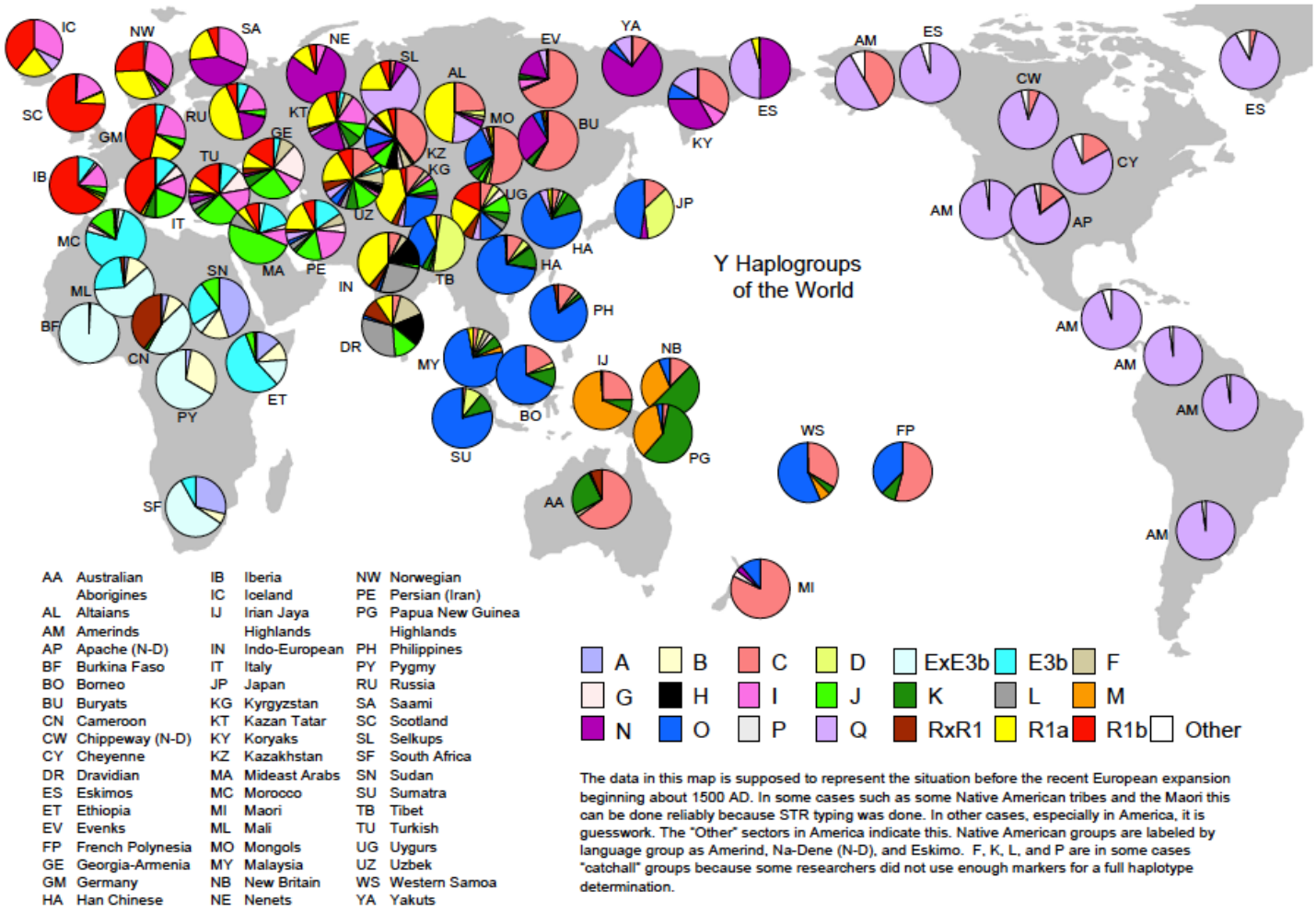



**Figure 5:** Worldwide NRY distributions. Figure taken from McDonald JD 2005. http://www.scs.illinois.edu/~mcdonald/WorldHaplogroupsMaps.pdf

Cheyenne) and central/southern Siberian populations (Kets, Yakut, Selkup and Altais) (Bortolini 2003). Overall the genetic data reveals a close relationship formed across the Bering Land Bridge, but can that data also reveal the number of migrations into the Americas?

Many population geneticists looking into the peopling of the Americas took the nearly uniform distribution of mtDNA and NRY haplogroups, as well as unique markers such as the 9RA to indicate a single migratory population that originally peopled North and South America. The genetic data showed only recent variation between coastal and

inland populations and practically no broad-scale distinctions between North, South, or Central American groups. A recent study of two of the small-scale haplogroups, D4h3 and X2a, broke down many of the original one-wave notions (Perergo 2009). The D4h3 haplogroup is unique to coastal populations, mainly in South America, but also across the U.S. coast and Central American coast. The X2a haplogroup is only found in the north-central plains region of Canada and the United states, right around the area where the ice-free corridor opened. Both haplogroups are dated between 16,700 and 15,500 bp, and overlapping divergence estimates reveal that D4h3a and X2a started spreading in America at approximately the same time. Their unique and isolated distributions, plus their pre-deglaciation dates make them prime indicators of the two-wave peopling of the Americas. The study received much support from some of the other major contributors to the debate (e.g., O'Rourke 2009). Identifying the genetic differences between one and two migrational populations—if both groups expanded from the same source population or two extremely similar source populations with recent ancestry—is nearly impossible. The previous information from Perergo (2009) is evidence of multiple migrations, but the haplotype differences could be the result of more recent genetic drift and not indicators of the ancestral populations. In order to shore-up many of the points of contention, the archaeological and linguist evidence throughout Siberia, Beringia, and the Americas play a major role in aiding the interpretation of the genetic data.

# Chapter 3: Methods

The Human Subjects Committee (HSC) and Institutional Review Board (IRB) of the University of Kansas approved this study. Participants for involvement in this research project gave verbal informed consent for the use of genetic materials for historical reconstruction.

## 3.1 Sample Collection

Blood samples and genealogic information used in this project were collected during field sessions conducted by the Laboratory of Biological Anthropology (LBA) throughout the 1990s and 2000s. Samples were utilized from the following populations: Altai (n=42), Chukchi (n=42), Koryak (n=26), and Aleut (n=37). Figure 6 provides the sampling locations of the populations. Samples for this study were screened for native heritage by selecting individuals who reported indigenous family history on both the maternal and paternal sides. .

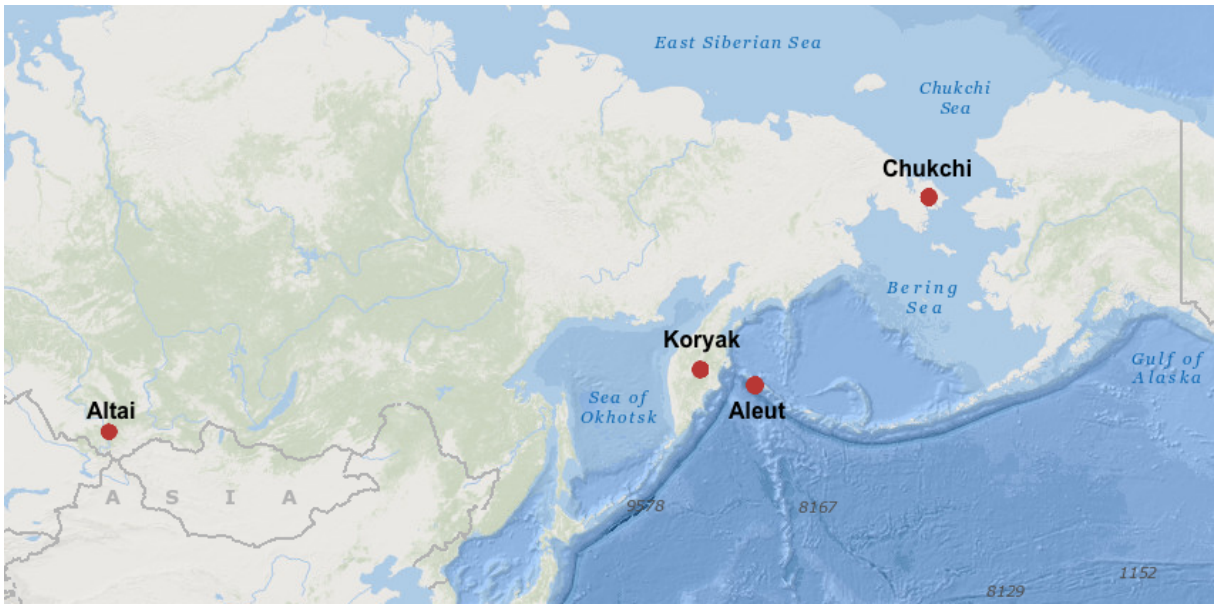



**Figure 6:** Map of Siberia showing the locations of sampled populations for this study. The map was made with ArcGIS.

### 3.2 Laboratory methods

DNA used for this study was all previously extracted from whole blood samples. PCR amplification was performed in a volume of 25 µL containing the following: 5.3 µL deionized $H_2O$, 3.0 µL 50mM $MgCl_2$, 1.0 µL BSA, 0.5 µL dNTPs, 5.0 µL 5x Buffer (Applied Biosystems, Faster City, CA), 0.2 µL Go Taq Flexi Polymerase (Applied Biosystems), 2.5 µL forward and reverse primers. Primers designed and used for this project ABO-1/2 and ABO-3/8, which amplify Exon 6 and Exon 7 of the ABO gene on chromosome 9, according to Ogasawara et al. (1996). Amplified DNA was analyzed via electrophoresis in a 2.0% agarose gel in order to verify its presence. The DNA was quantified by spectrophotometry with a NanoDrop according to the manufacturer's protocol for DNA (Thermo Fisher Scientific Inc.). Quantified DNA was then diluted with deionized water to 15-25 ng/µL and sent to Beckman Coulter Genomics for Sangor Sequencing. The G542A mutation was then confirmed by RFLP digestion with NheI restriction enzyme (Olsson et al. 1998).

### 3.3 Comparative Population Data

Allelic frequencies for 13 North, Mesoamerican, and South American populations were compiled from the literature (Olsson et al. 1998, Roubinet et al. 2001, Barjas-Castro et al. 2003, Llop et al 2006, Estrada-Mena et al 2010, Villanea et al. 2013). These were the only Native American groups that have been specifically tested for the G542A mutation leading to the O1v542 haplotype. The 13 populations were used in all the between population analyses.

### 3.4 Analytical Methods

*3.4.1 Sequence Analysis and haplotyping*

DNA sequences were aligned and inspected for reliability and heterozygosity using Sequencher 5.2 (Gene Codes Corporation, Ann Arbor, MI). Exons 6 and 7 were combined into an 826 base pair dataset for each sample and compared to the Cambridge Reference Sequence. Haplotypes were assigned according to Ogasawara et al. (1996) and Estrada-Mena et al. (2010).

A median joining network was constructed using the Network software package in order to show the mutational connection between the ABO haplotypes in the current study (Bandelt et al. 1999).

*3.4.2 Within Population Diversity*

Haplotype frequency and gene and nucleotide diversity were calculated using Arlequin version 3.5.1.2 (Excoffier et al. 2005; Excoffier and Lischer 2010) for all seventeen populations. These values provide information on the amount of variation

within the population, and inferences of migratory patterns can be made from the values (Nei 1987; Nei and Li 1979; Tajima 1983). Gene and nucleotide diversity measure the probability of differences occurring in two randomly chosen haplotypes or base pairs, respectively. Stochastic effects are mediated by gene diversity calculations since they examine haplotypes, which are accumulations of multiple single base pair mutations (Helgason et al. 2003; Nicholson et al. 2002).

### *3.4.3 Between Population Diversity*

Multi-dimensional Scaling (MDS), Analysis of Molecular Variance (AMOVA), Neighbor-Joining Trees (NJ), and Mantel tests were all computed and constructed in order to assess the relationship between the populations from this study and the comparative populations from previous publications. Because of different substitution rates between transversions (T or C ← → A or G) and transitions (T←→C or A←→G) based on the number of hydrogen bonding sites, the Kimura 2-paramater (K2P) model was implemented (Kimura 1980). The K2P distance model was used for all within and among population analyses to establish reliable diversity measures and distance matrices for plotting connections.

#### *3.4.3.1 MDS*

Multi-dimensional scaling utilizes the pairwise differences created with the K2P model to infer a diagram of relationships among populations (Torgerson 1952; Kruskal 1964a; Kruskal 1964b). A geometric shape can be used to represent the relationships when one, two, or three dimensions best incorporates the greatest amount of variation. A stress test was first performed to ascertain the optimal dimensionality that could fully

represent the data. The lower the stress value, the less the distances had to be altered in order to fit the plot. Stress values below 0.1 are necessary to justify the results, while less than 0.05 is preferential (Kruskal and Wish 1978). The choice of dimensionality was then made to keep the number of dimensions low, so a geometric representation could be constructed, and also to keep the stress value as low as possible. After establishing dimensionality, the MDS plot was constructed for all 17 populations in NTSYSpc version 2.2 (Rohlf 2008).

*3.4.3.2 AMOVA*

Analysis of Molecular Variance was used in order to establish the amount of variation within *versus* between population groups. AMOVA uses the same framework as the more common Analysis of Variance (ANOVA) method (Cockerham 1969, 1973). The hierarchical analysis works by partitioning total variance into three covariant components: intra-population, inter-population within groups, and inter-group differences (Excoffier et al. 1992). The algorithm established by Excoffier (2000) calculates F-statistic values that reflect the proportion of haplotypic variance attributed to the covariant components. In an AMOVA calculation, FST, FSC, and FCT values are calculated for variance among populations relative to the total variance, variance among populations within groups, and variance among groups relative to total variance.

Researcher-assigned groups are required to calculate the statistic. The 17 populations were divided into four groups based on location and known mtDNA and NRY haplogroups: Siberian (Altai, Chukchi and Koryak), North American (Aleut, Tlingit and Olhone), Mesoamerican (Nahua, Mazahua and Maya), and South American (Cayapa, Arara, Aymara Bolivia, Aymara Chile, Huilliche, Yanomama, Parakana and Kayapo).

*3.4.3.3 NJ*

Neighbor Joining trees are established like MDS plots from distance data. The clustering method allows creation of phylogenetic trees based on closeness of taxa according to their evolutionary divergence times. Relationships and divergences can then be inferred according to proximity between nodes (Saito and Nei 1987). For this study, the data were bootstrapped 1000 times in order to establish the most representative tree. The first tree was constructed using MEGA version 6.06 (Tamura et al. 2007; 2011), while a second tree was constructed according to the Unweighted Pair-Group Method with Arithmetic mean (UPGMA) (Sokal and Michener 1958). The UPGMA method assumes all lineages evolve at the same rate, so branch lengths within nodes do not vary, but more refined trees are created. In the UPGMA, distances among all individuals are averaged to create a mean distance between populations. This method establishes quantized separation between groups based on the least dissimilar distances, and provides a rooted tree with the root at an equal distance from all populations.

A combination of the branching pattern from the UPGMA and the distances from the NJ tree was generated for a more accurate estimation of relationships when assuming genetic drift links the populations. This combination is effective for the ABO blood group data because there are differences in the number of mutations between haplogroups, favoring NJ. However, when assuming a migratory pattern where the various haplogroups were transferred by founder effect, the mutational separation is significant only for newly established haplogroups. This favors UPGMA since the only "new" haplogroups (Ov7, O32, O33) observed, that are not present in the Siberian

populations, are the result of a single mutational step and may have been brought by later admixture. The combination tree was created in Adobe Illustrator CS6.

*3.4.3.4 Mantel Test*

A mantel test was performed by comparing the pairwise K2P distance matrix with a geographic distance matrix. A Microsoft Excel add-on, GenAlX 6.5, was used to perform the analysis (Peakall and Smouse 2012). GenAlX 6.5 was also used to create the geographic distances by inputting longitude and latitude values that were then adjusted for the spherical shape of the earth when establishing direct distances between points. The points between matrices were plotted on an XY graph and tested with a goodness of fit line to determine the $R^2$ value.

*3.4.4 Gene Flow and Back Migrations*

To estimate the amount of gene flow necessary to explain the frequency of the O1v542 allele in the sample populations, a simple statistic was adapted from Bernstein (1931) and used according to Schroeder et al. (2007). This estimate will determine if back-migration from the Americas is a reasonable explanation for the presence of O1v542 by determining the amount of gene flow (m) required. The calculation is represented by:

$$m = (p_h - p_2)/(p_1 - p_2) \qquad \text{(Equation 1)}$$

Where $p_h$ is the observed average frequency of O1v542 in the population being tested, p1 is the observed average frequency of O1v542 in all other Native American

populations, and p2 is the frequency of the O1v542 allele in the population being tested prior to gene flow. The resulting formula is:

$$m = (p_h)/(p_1) \qquad \text{(Equation 2)}$$

Since there is the assumption that the populations being tested did not have the marker being examined until back migrations occurred ($p_2 = 0$).

# Chapter 4: Results

## 4.1 Within Population Diversity

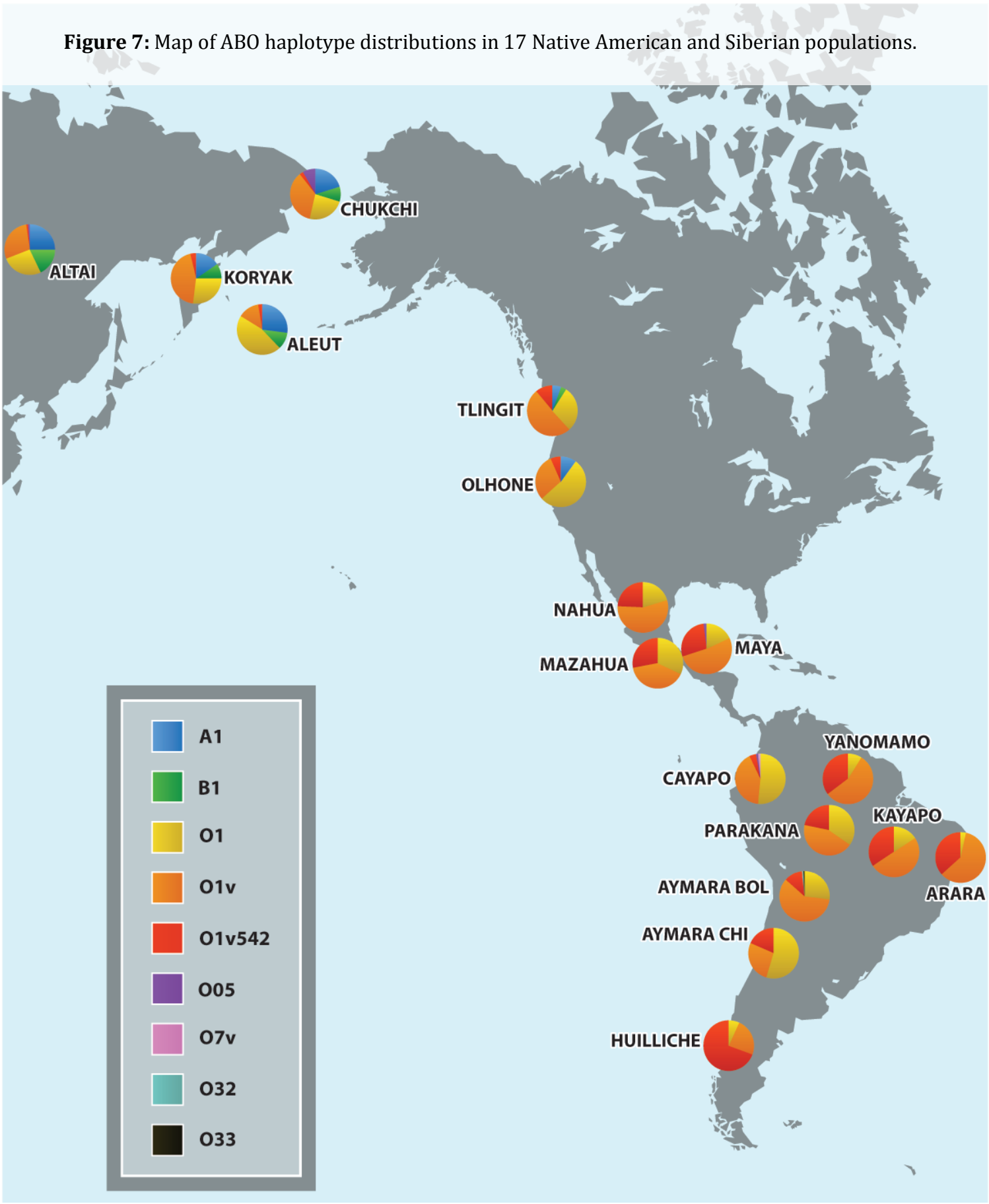
Haplotypes and frequencies of O1v542 for all samples used in this study and those referenced from previous studies are compiled in Table 2. Haplotypes were assessed according to the mutational differences shown previously in Table 1. All haplotypes were initially predicted based on Exon 7 sequences and confirmed with the corresponding Exon 6 sequences. The necessity of sequencing each sample twice combined with the 100% confirmation of haplotypes for all samples provide high confidence in the reliability of the determined haplotypes. The G542A mutation was more difficult to identify by separate sequencing because a baseline adenine signal appeared on most sequences, so the RFLP provided support to its identification. Again, 100% of sequences that demonstrated higher than normal adenine 542 peaks matched with 100% of the restricted sequences, validating results. A geographic map representing all populations and their ABO haplotype frequencies can be seen in Figure 7.

The median joining network indicates the stepwise differences between ABO alleles (Figure 8). The size of the nodes in the network corresponds to the frequency in the 17 Native American and Siberian populations.

*Table 2. Sample size and haplotypes used for analysis. Samples represent all Native American groups tested for O1v542 at the time of this study.*

| Population | Sample Size | Total Chrom. | A1 | B1 | O1 | O1v | O1v542 | O05 | Ov7 | O32 | O33 | O1v542 Frequency | Reference |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Aleut | 37 | 74 | 20 | 8 | 34 | 10 | 2 | - | - | - | - | 0.03 | This Study |
| Altai | 42 | 84 | 21 | 15 | 22 | 24 | 1 | 1 | - | - | - | 0.01 | This Study |
| Arara | 15 | 30 | - | - | 1 | 18 | 11 | - | - | - | - | 0.38 | Olsson et al. 1998 |
| Aymara Bolivia | 63 | 126 | - | - | 34 | 75 | 15 | - | - | 1 | 1 | 0.12 | Estrada-Mena et al. 2010 |
| Aymara Chile | 84 | 168 | - | - | 59 | 89 | 20 | - | - | - | - | 0.12 | Llop et al. 2006 |
| Cayapa | 35 | 70 | - | - | 36 | 29 | 3 | 1 | 1 | - | - | 0.04 | Estrada-Mena et al. 2010 |
| Chukchi | 42 | 84 | 17 | 8 | 20 | 30 | 2 | 7 | - | - | - | 0.02 | This Study |
| Huilliche | 67 | 134 | - | - | 9 | 32 | 93 | - | - | - | - | 0.69 | Llop et al. 2006 |
| Kayapo | 16 | 32 | - | - | 5 | 16 | 11 | - | - | - | - | 0.33 | Olsson et al. 1998 |
| Koryak | 26 | 52 | 8 | 5 | 14 | 23 | 2 | - | - | - | - | 0.04 | This Study |
| Maya | 50 | 100 | - | - | 18 | 52 | 28 | 2 | - | - | - | 0.28 | Estrada-Mena et al. 2010 |
| Mazahua | 50 | 100 | - | - | 32 | 40 | 28 | - | - | - | - | 0.28 | Estrada-Mena et al. 2010 |
| Nahua | 37 | 74 | - | - | 15 | 41 | 18 | - | - | - | - | 0.24 | Estrada-Mena et al. 2010 |
| Ohlone | 15 | 30 | 3 | - | 16 | 9 | 2 | - | - | - | - | 0.07 | Villanea et al. 2013 |
| Parakana | 62 | 124 | - | - | 43 | 54 | 27 | - | - | - | - | 0.22 | Barjas-Castro et al. 2003 |
| Tlingit | 72 | 144 | 8 | 5 | 42 | 73 | 16 | - | - | - | - | 0.11 | Villanea et al. 2013 |
| Yanomama | 17 | 34 | - | - | 3 | 19 | 12 | - | - | - | - | 0.36 | Olsson et al. 1998 |

**Figure 7:** Map of ABO haplotype distributions in 17 Native American and Siberian populations.



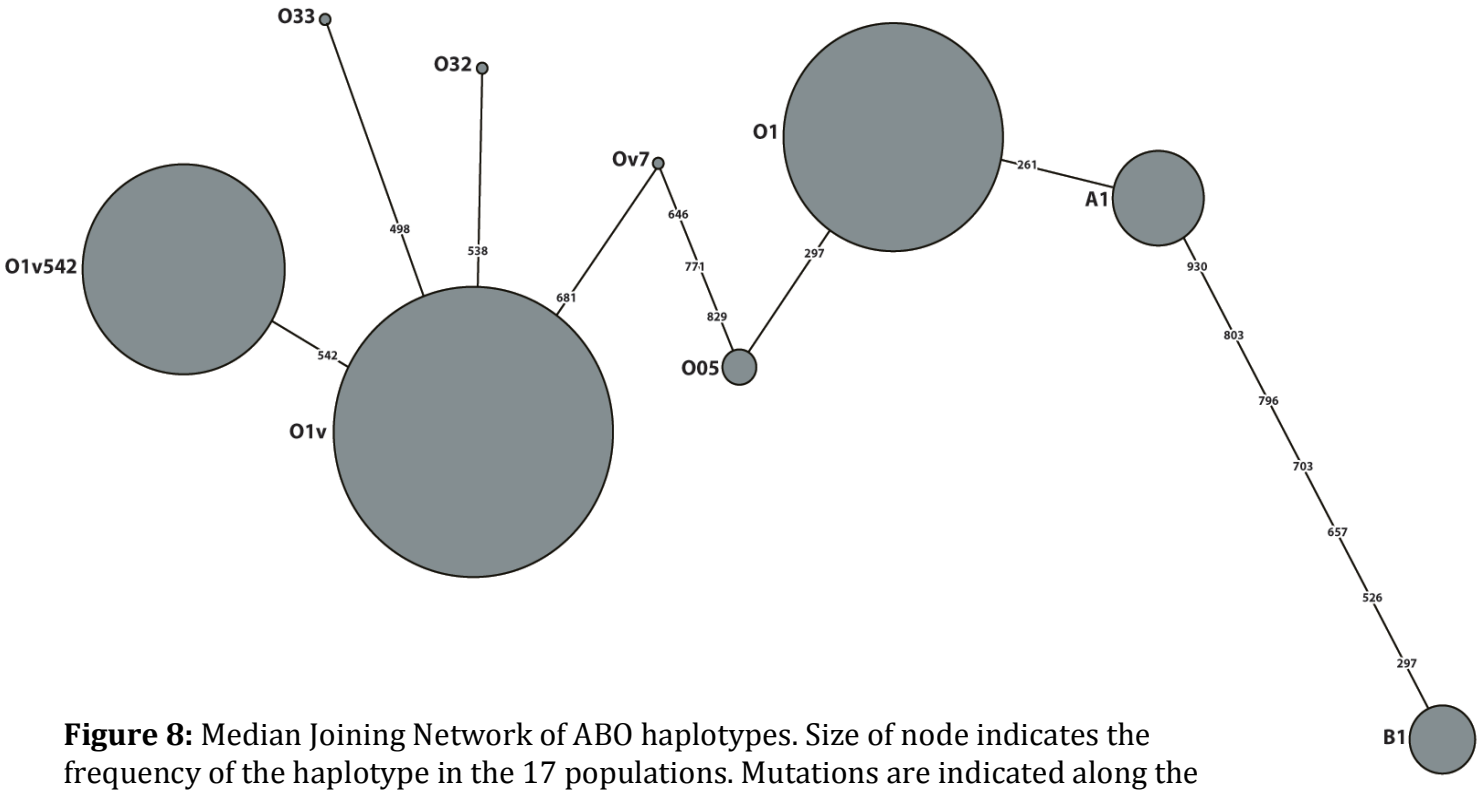



**Figure 8:** Median Joining Network of ABO haplotypes. Size of node indicates the frequency of the haplotype in the 17 populations. Mutations are indicated along the lines that connect nodes.

Nucleotide and gene diversity measurements were computed for all 17 populations in order to establish within population diversity (Table 3). The Huilliche and Arara had the lowest gene and nucleotide diversity measurements, while the four populations specific to this study (Aleut, Altai, Chukchi and Koryak) had the highest gene and nucleotide diversity.

*Table 3. Gene and Nucleotide Diversity measurements.*

| Population | Gene Diversity (S.D.) | Nucleotide Diversity (S.D.) |
|---|---|---|
| **Aleut** | **0.732 (0.04)** | **0.0042 (0.002)** |
| **Altai** | **0.780 (0.02)** | **0.0059 (0.003)** |
| Arara | 0.522 (0.06) | 0.0010 (0.001) |
| AymaraBol | 0.563 (0.03) | 0.0027 (0.002) |
| AymaraChi | 0.585 (0.02) | 0.0030 (0.002) |
| Cayapa | 0.570 (0.03) | 0.0032 (0.002) |
| **Chukchi** | **0.776 (0.02)** | **0.0051 (0.003)** |
| Huilliche | 0.460 (0.04) | 0.0013 (0.001) |
| Kayapo | 0.627 (0.05) | 0.0022 (0.002) |
| **Koryak** | **0.720 (0.04)** | **0.0052 (0.003)** |
| Maya | 0.625 (0.03) | 0.0024 (0.002) |
| Mazahua | 0.670 (0.01) | 0.0032 (0.002) |
| Nahua | 0.601 (0.04) | 0.0024 (0.002) |
| Ohlone | 0.632 (0.07) | 0.0035 (0.002) |
| Parakana | 0.648 (0.02) | 0.0032 (0.002) |
| Tlingit | 0.646 (0.03) | 0.0039 (0.002) |
| Yanomama | 0.572 (0.05) | 0.0016 (0.001) |

### 4.2 Between Population Diversity

The individual sequences were separated into their population groups and the Kimura-2P distance matrix was computed (Table 4). P-values of the established distances are recorded above the * diagonal in Table 4. This matrix was used for all the between population diversity statistics.

*Table 4. Pairwise nucleotide diversity using Kimura 2P model among 17 Native American and Siberian populations. P-values are given above the diagonal*

| | Chukchi | Altai | Koryak | Aleut | Tlingit | Ohlone | Nahua | Mazahua | Maya | AymaraBol | Cayapa | AymaraChi | Huilliche | Parakana | Kayapo | Arara | Yanomama |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chukchi | * | 0.155 | 0.501 | 0.002 | 0.000 | 0.196 | 0.000 | 0.000 | 0.000 | 0.000 | 0.013 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Altai | 0.010 | * | 0.026 | 0.029 | 0.000 | 0.015 | 0.000 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Koryak | -0.005 | 0.033 | * | 0.000 | 0.043 | 0.146 | 0.000 | 0.004 | 0.000 | 0.000 | 0.120 | 0.003 | 0.000 | 0.009 | 0.000 | 0.000 | 0.000 |
| Aleut | 0.062 | 0.026 | 0.114 | * | 0.000 | 0.009 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Tlingit | 0.074 | 0.152 | 0.026 | 0.253 | * | 0.008 | 0.009 | 0.086 | 0.001 | 0.034 | 0.024 | 0.318 | 0.000 | 0.213 | 0.007 | 0.000 | 0.000 |
| Ohlone | 0.014 | 0.057 | 0.022 | 0.076 | 0.085 | * | 0.000 | 0.001 | 0.000 | 0.000 | 0.291 | 0.002 | 0.000 | 0.002 | 0.000 | 0.000 | 0.000 |
| Nahua | 0.230 | 0.307 | 0.165 | 0.448 | 0.053 | 0.293 | * | 0.091 | 0.912 | 0.178 | 0.000 | 0.009 | 0.000 | 0.032 | 0.563 | 0.022 | 0.131 |
| Mazahua | 0.145 | 0.229 | 0.088 | 0.342 | 0.012 | 0.157 | 0.018 | * | 0.048 | 0.173 | 0.001 | 0.254 | 0.000 | 0.658 | 0.108 | 0.000 | 0.011 |
| Maya | 0.248 | 0.328 | 0.183 | 0.464 | 0.064 | 0.309 | -0.011 | 0.023 | * | 0.069 | 0.000 | 0.003 | 0.000 | 0.017 | 0.647 | 0.043 | 0.188 |
| AymaraBol | 0.181 | 0.271 | 0.115 | 0.396 | 0.019 | 0.212 | 0.007 | 0.007 | 0.016 | * | 0.000 | 0.168 | 0.000 | 0.155 | 0.104 | 0.002 | 0.010 |
| Cayapa | 0.039 | 0.109 | 0.022 | 0.169 | 0.035 | 0.001 | 0.190 | 0.085 | 0.206 | 0.117 | * | 0.014 | 0.000 | 0.008 | 0.000 | 0.000 | 0.000 |
| AymaraChi | 0.124 | 0.217 | 0.067 | 0.321 | 0.000 | 0.125 | 0.042 | 0.004 | 0.052 | 0.007 | 0.052 | * | 0.000 | 0.503 | 0.014 | 0.000 | 0.000 |
| Huilliche | 0.472 | 0.525 | 0.437 | 0.655 | 0.277 | 0.593 | 0.167 | 0.217 | 0.142 | 0.244 | 0.481 | 0.279 | * | 0.000 | 0.004 | 0.004 | 0.005 |
| Parakana | 0.126 | 0.213 | 0.071 | 0.319 | 0.003 | 0.128 | 0.032 | -0.006 | 0.040 | 0.008 | 0.060 | -0.003 | 0.250 | * | 0.048 | 0.000 | 0.001 |
| Kayapo | 0.249 | 0.314 | 0.185 | 0.466 | 0.081 | 0.338 | -0.012 | 0.035 | -0.015 | 0.033 | 0.236 | 0.076 | 0.104 | 0.057 | * | 0.242 | 0.619 |
| Arara | 0.356 | 0.408 | 0.302 | 0.570 | 0.182 | 0.510 | 0.067 | 0.144 | 0.056 | 0.133 | 0.381 | 0.188 | 0.083 | 0.170 | 0.024 | * | 0.704 |
| Yanomama | 0.311 | 0.370 | 0.252 | 0.528 | 0.135 | 0.436 | 0.021 | 0.091 | 0.015 | 0.082 | 0.318 | 0.135 | 0.081 | 0.116 | -0.016 | -0.017 | * |

*4.2.1 MDS*

The 2-dimensional Multi-Dimensional Scaling (MDS) plot constructed in NTSYS ver 2.2 (Rholf 2008) was most representative of the data while still being geometrically relevant according to the stress test (Figure 9). The stress was 0.0217, which falls well below the 0.1 value that is required to substantiate a resulting plot. The four groups that were established for the Analysis of Molecular Variance (AMOVA) are identified in Figure 10 to show any regional clustering.

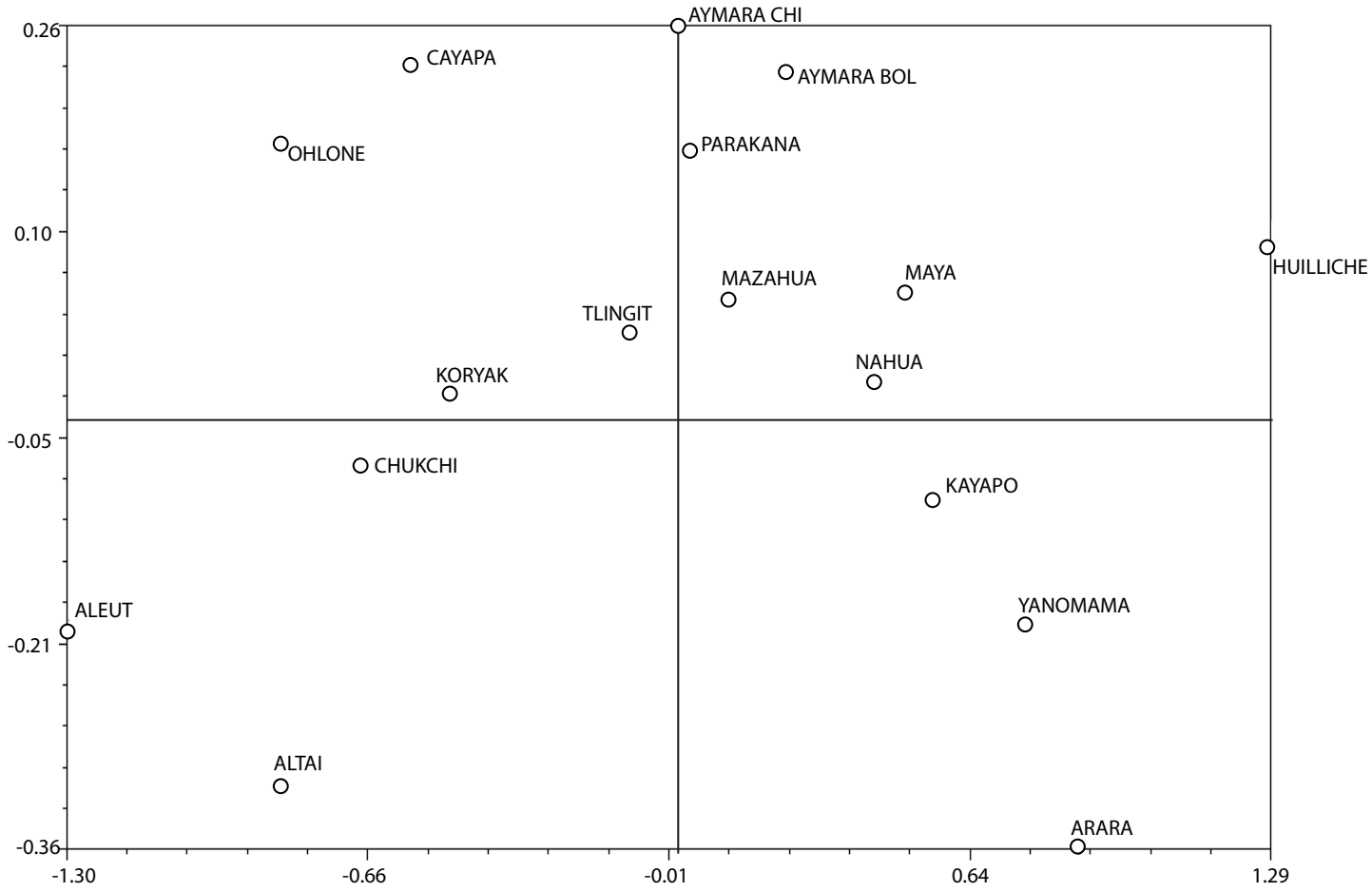



**Figure 9:** Multi Dimensional Scaling plot of 17 Native American and Siberian Populations. Distances were established using the Kimura 2-Parameter method.

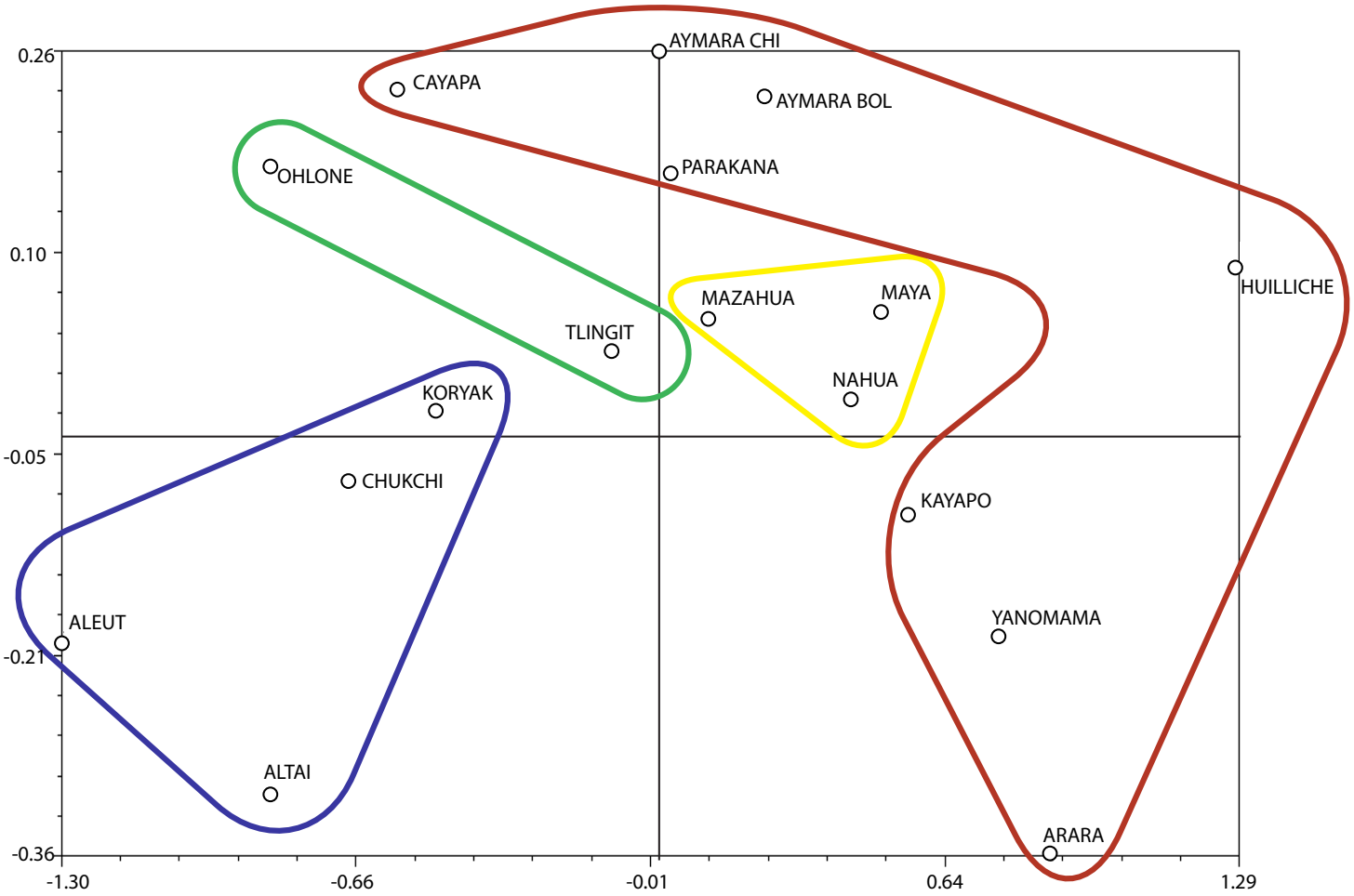



**Figure 10:** Multi Dimensional Scaling plot with regional group clusters identified within circles. Blue is the Siberian group, green is the North American group, yellow is the Mesoamerican group and red is the South American group.

*4.2.2 AMOVA*

Analysis of Molecular Variance (AMOVA) was used by segregating the 17 populations into four regional groups: Siberian, North American, Meso American, and South American. The statistic computes relative percentages of diversity within a population, between populations within a group, and between groups (Table 5). The majority (80.85%) of the total diversity was within populations. There was about 10% diversity both among populations within groups and among groups. The fixation indices agree with the % variance findings, and all were significant to at least 95% confidence level.

*Table 5. Analysis of Molecular Variance based on ABO Exon 6 and 7 sequences with populations grouped according to geographic region*

| Source of Variation | d.f. | SSD | Variance | % Total Variance | Fixation Index | P-Values |
|---|---|---|---|---|---|---|
| Among Groups | 3 | 189.1 | 0.144 | 9.36 | FCT = 0.094 | 0.01466 |
| Among Populations Within Groups | 13 | 176.4 | 0.151 | 9.79 | FSC = 0.108 | <0.00001 |
| Within Populations | 1443 | 1797.8 | 1.249 | 80.85 | FST = 0.192 | <0.00001 |
| **Total** | 1459 | 2163.8 | 1.541 | | | |

*4.2.3 NJ*

A Neighbor-Joining tree (NJ) was computed utilizing the K2P distances in MEGA version 6.06 (Tamura et al. 2007; 2011) (Figure 11). A second tree was computed with the Unweighted Pair-Group Method with Arithmetic Mean (UPGMA), which standardizes evolution rates and provides a better-arranged tree (Figure 12). The UPGMA branching is slightly different than that of the NJ tree and provides better grouping because the seminal hypothesis demonstrated by mtDNA, NRY, and autosomal markers is that the populations resulted from 1-3 migrations or back-migrations. The two trees were combined in Adobe Illustrator CS6, utilizing UPGMA relationships and NJ distances (Figure 13).

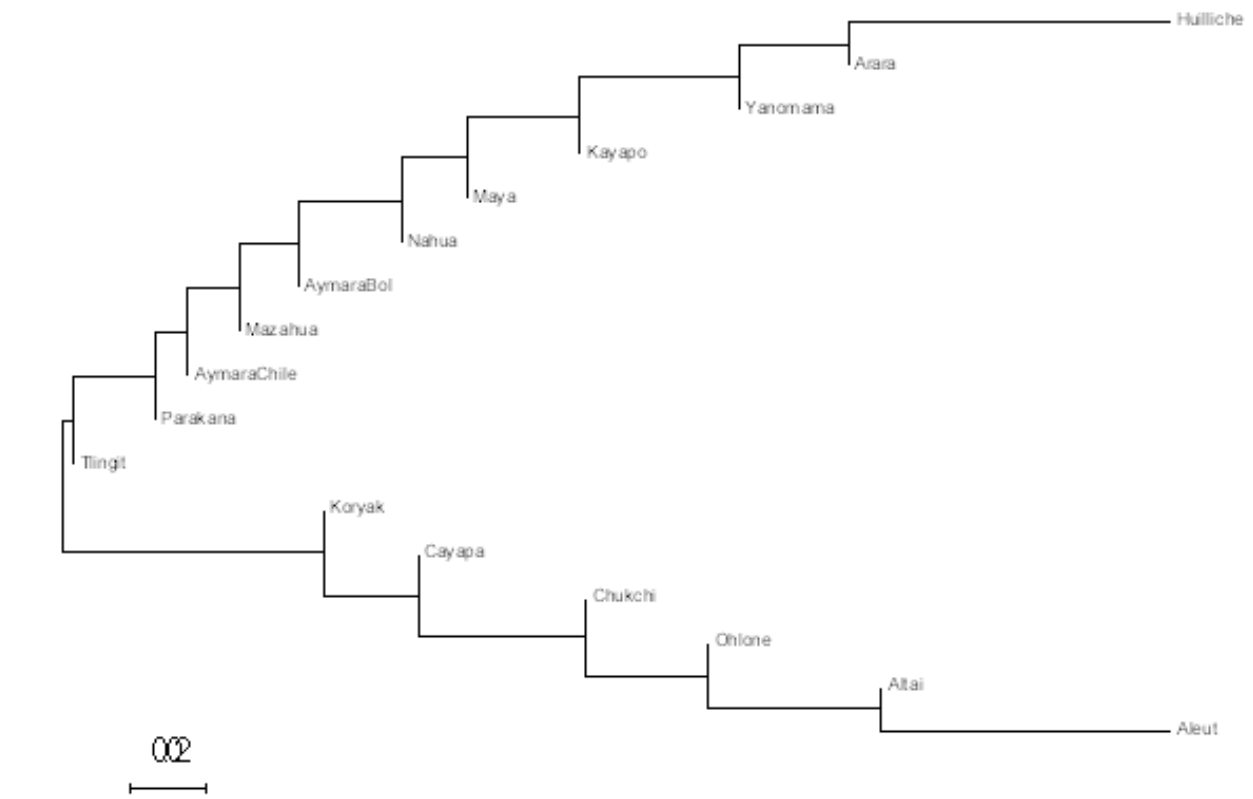



**Figure 11:** NJ tree of 17 Native American and Siberian populations using K2P distances.

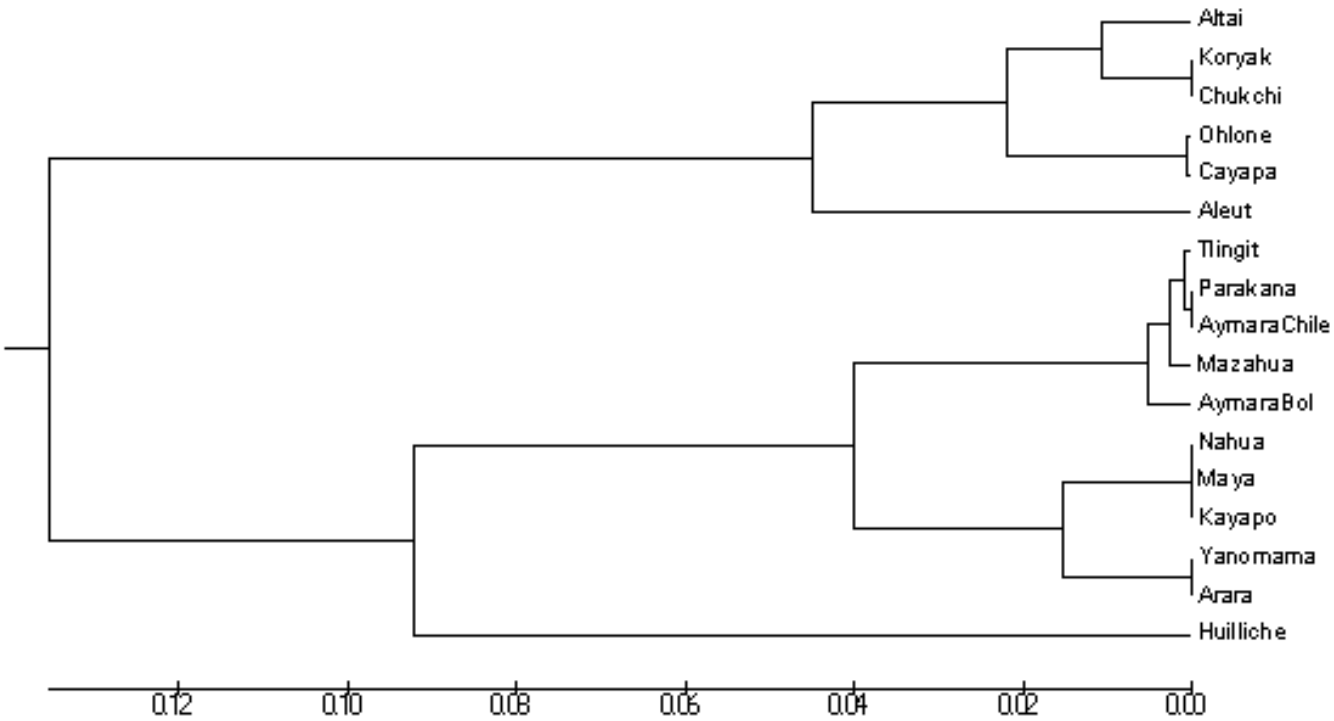



**Figure 12:** Rooted UPGMA tree of 17 Native American and Siberian populations.

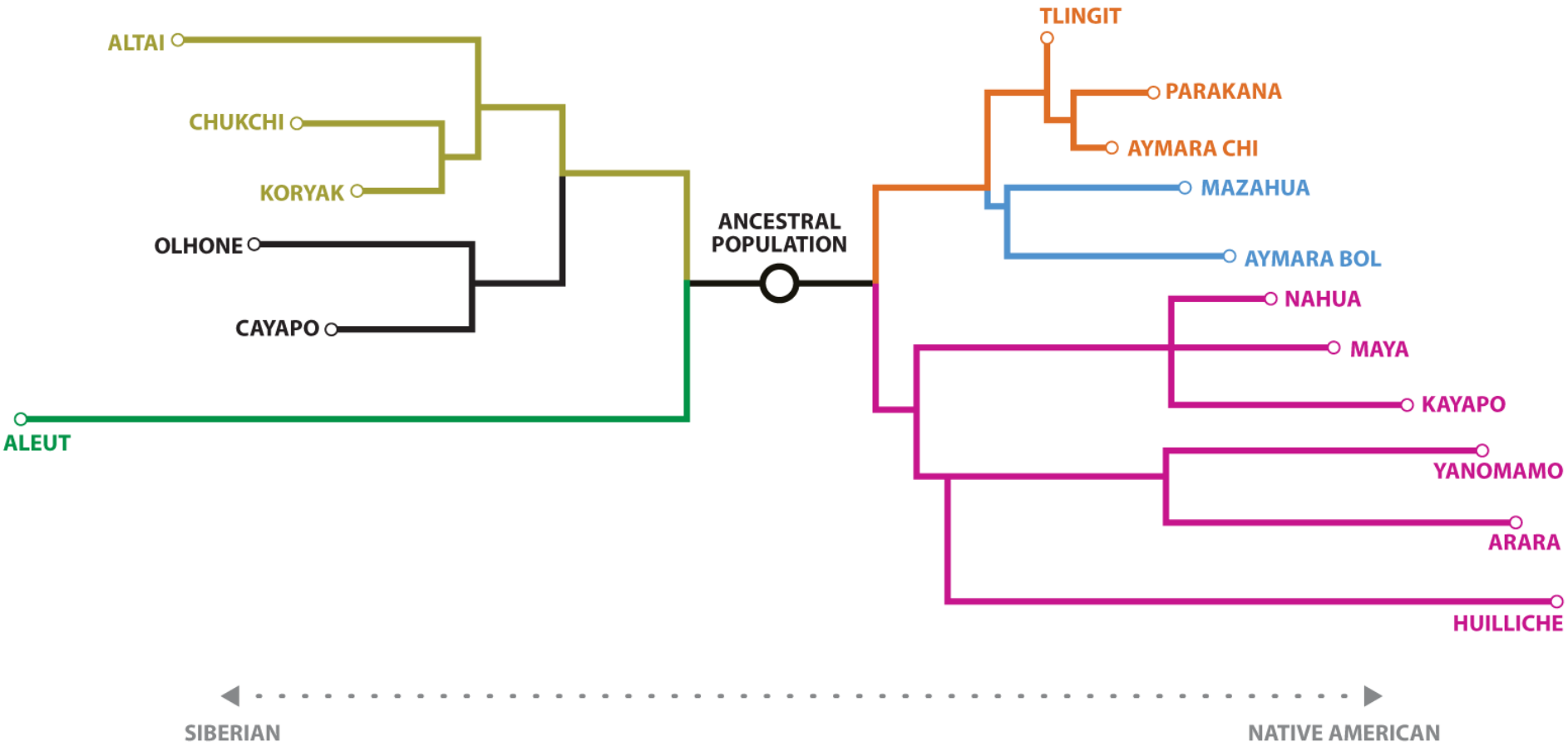
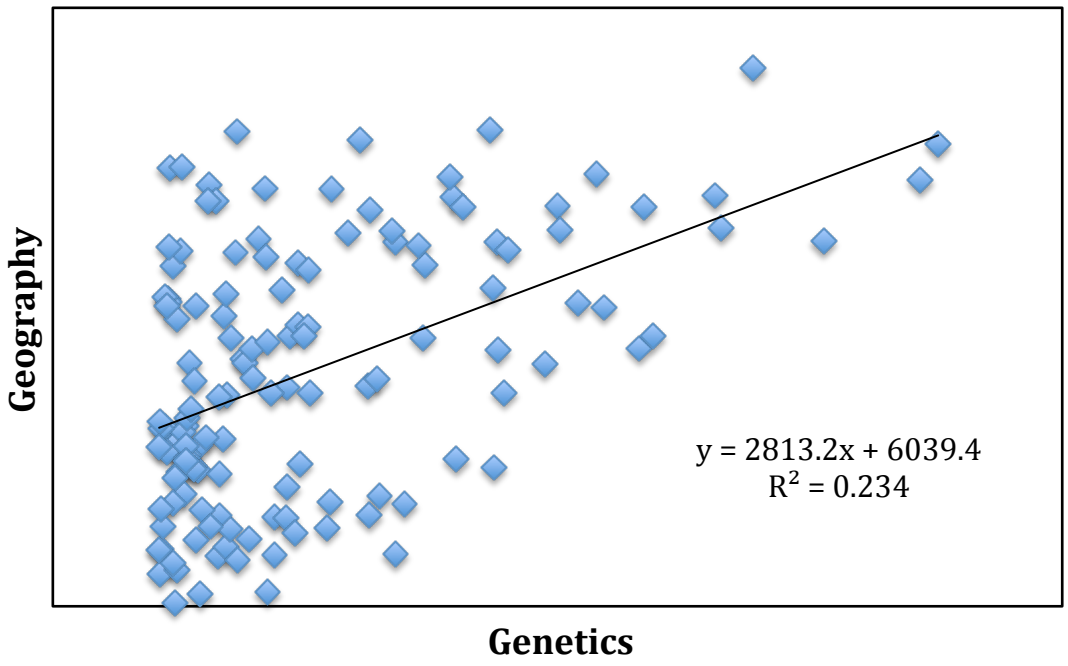



**Figure 13:** Combination of UPGMA and NJ trees. The tree is rooted using the most recent common ancestor between the Siberian and Native American branches. Different color paths represent potential radial migrations from the ancestral Beringian population according to a multi-wave expansion hypothesis.

*4.2.4 Mantel Test*

In order to test the relationship between genetics and geography for the ABO blood group region for the 17 populations, a Mantel Randomization test was performed. The K2P distance matrix was compared to a geographic distance matrix formed from longitude and latitude values for the sampled populations (Figure 14). The results revealed a correlation of 0.234, which is not significant.

**Figure 14:** Mantel test of geographic vs genetic distances with goodness of fit line.



### 4.3 Gene Flow and Back Migration

Using the formula from Equation 1, the amount of gene flow ($m$) necessary to result in the frequency of O1v542 observed in the Siberian and Aleut populations was calculated (Shroeder et al. 2007), where $p_h$ is the frequency of O1v54 in the population being tested, while $p_1$ is the total frequency of O1v542 in all populations further along the migratory path or in the Americas. For example, the Altai used all of the other 16 populations to determine $p_1$, while the Aleut used only the 13 Native American populations. The $m$ min and max were calculated using the population with the highest and lowest frequency of O1v542. For all four populations, the $m$ max happened to be calculated with the closest geographic population, indicating the amount of gene flow necessary, from the most accessible source.

*Table 7: The amount of gene flow (m) necessary to explain the frequency of O1v542. $p_h$ is the frequency of O1v542 in the population, $p_1$ is the frequency of O1v542 in all other populations.*

| Population | $p_h$ | $p_1$ | $m$ | $m$ min | $m$ max |
|---|---|---|---|---|---|
| Aleut | 0.0270 | 0.249 | 10.8% | 3.9% | 67.6% |
| Altai | 0.0119 | 0.208 | 5.7% | 1.7% | 59.5% |
| Chukchi | 0.0238 | 0.234 | 10.2% | 3.5% | 79.4% |
| Koryak | 0.0385 | 0.233 | 16.5% | 5.6% | 192.3% |

# Chapter 5: Discussion

The following sections interpret and present the previous results by providing a narrative of exploration and expansion into a previously uninhabited continent.

## 5.1 Within Population Diversity

The frequency of the O1v542 mutation in Native Americans ranges from 4% to 70% (Villanea et al. 2013). Of the four populations used in this study, only the Koryak have a percentage that falls in the known range (4%), while the Altai, Chukchi and Aleut have slightly lower frequencies: 1%, 2%, and 3%, respectively. The raw number of individuals with one copy of the allele (none were homozygous for O1v542) was 1 for the Altai and 2 for each of the other three populations. The A blood group was only present in the North American and Siberian populations, with the four populations from this study having a much higher frequency. This supports previous findings that indicate a southward increase in frequency of the O blood group in the Americas (Yamamoto 2000). All populations had the three O haplotypes: O1, O1v and O1v542. Haplotype O1v was the most prominent as is indicated in the Median-Joining Network. The network shows the O05 haplotype branching from O1 and the other three haplotypes (Ov7, O32 and O33) splitting from O1v by a single mutational difference.

Gene and nucleotide diversity values further explain the possible expansion beyond basic frequencies. It is expected that diversity values would decrease if founder events were occurring, as they would during multiple migrations. This hypothesis is substantiated since the Altai have the greatest diversity, followed by the Chukchi, Aleut

and Koryak. The high diversity in the Aleut is most likely the result of recent admixture from European populations. The genetic diversity on Bering Island, the site where the Aleut samples were taken, is the direct result of a unique founder event. Russian soldiers uprooted Aleuts from other islands and relocated them to Bering Island. The Island's governor then encouraged Russian men to admix with Aleut females as a means to prevent tribal inbreeding (Crawford 2007b). The pattern may have led to greater diversity than would have been observed in the original Aleut population, although all individuals used in this study reported having no known recent Russian ancestry.

Gene and nucleotide diversity continue to decrease as the populations move further south into the Americas. The first-nations group, Tlingit, and the Northern United States group, Ohlone, both have high values. The Maya, Mazahua, and Nahua — Meso American groups — also have the next highest values. The Parakana have values around those of the Meso-American groups, which is not surprising since they are located in north-central South America. The Arara and Huilliche have the lowest diversity values and are both the most geographically isolated South American groups. The Kayapo's diversity values are higher than expected for their geographic location. Two possible explanations are: (a) recent admixture from other populations or (b) they were quickly populated during an expansion from the Meso-American or Parakana groups. Other analyses may provide further indication of which explanation is best.

### 5.2 Between Population Diversity

Between population diversity is an effective gauge of population relationships, which can then be extrapolated into historical connectivity. Various two-dimensional

plotting statistics (MDS and NJ) provide visual representations of those relationships, while other computational analyses (AMOVA and Mantel Tests) give values that tell how the diversity is characterized.

### *5.2.1 MDS*

The low stress value (0.0217) strengthens the reliability of the MDS plot. Figure 10, where the clusters are identified, shows separation between the four regional groups with a few outliers. The Koryak group clusters closer to the Native American groups, particularly the Tlingit and the Mesoamerican populations. The Aleut fall far away from the other Native American populations, supporting the three-wave model, which places them as the last migratory wave well over 10,000 years after the initial wave. That time separation would result in large genetic differentiation. The Chukchi fall in-between the Altai and Koryak but closer to the Koryak, meaning the Chukchi and Koryak could both be source populations or both derivatives of the same source population that populated the Americas.

The Ohlone and Cayapa are outliers from their expected positions. The Ohlone samples were taken from an ancient burial site radiocarbon dated between 250 and 2200 bp, but the burial culture is associated with the Late Period (750-230 bp) (Villanea et al 2013). The dates associated with this group mean that it was either a much later migration from the Beringia region, had significant admixture from another population, or that it is more representative of ancient Native Americans and the other groups are outliers from that perspective. Since only forty-one of the sixty-eight sampled individuals (~60%) belonged to one of the four known Native American haplogroups (A,

B, C or D), it is most likely that there was admixture from an outside population. This hypothesis is supported by previous research that showed >40% European admixture in US Native Americans (Hunley and Healy 2011). The Cayapa, likewise, are suspected of having significant European admixture (Estrada-Mena et al. 2010).

The other groups cluster in a trajectory similar to the paths possibly taken throughout the Americas. The east coast Native American populations (Parakana, Aymara Chile, and Aymara Boliva) cluster to one side of the Meso-American group, but the west coast Native American populations (Kayapo, Yanomama, and Arara) all split to the other side of the Meso- American group. The Huilliche are geographically distinct from the other South American populations since they are much further south and more central. The fact that they are isolated on the x-axis of the MDS, but in-between the east and west coast South American populations on the y-axis, aligns well with their geographic proximity.

### *5.2.2 AMOVA*

The 17 populations were segregated into their four geographic groups in order to test the percent of the genetic diversity within *versus* between populations and groups. The results indicated that the majority (80.85%) of the variation occurs within the individual populations. ~10% of the variation occurs each between populations within groups and between groups. These values correspond to many known human population studies (Barbujani et al. 1997; Stone and Stoneking 1998).

*5.2.3 NJ*

The Neighbor-Joining tree and the UPGMA tree both indicate a fundamental split between two groups: the Altai, Aleut, Chukchi, Koryak, Ohlone and Cayapa populations on one side and all of the other Native American populations on the other. The Altai represent the root of the Siberian and outliers group while the Tlingit represent the root of the Native American group. The Chukchi and Koryak both split from the Altai with the Koryak population being closer to the Native American groups, which was also shown in the MDS plot. The Cayapa and Ohlone remain outliers, but further strengthen the admixture argument since they appear to fall closer to the Siberian side of the tree than the Native American. The Aleuts form their own more distant branch from the other Siberian populations, still pointing to them being a later migratory group to the Americas. This could also indicate a substantial amount of recent admixture, but this possibility was minimized by selecting individuals who claimed native ancestry on both family lines.

The Native American side of the tree suggests three migrations in the Americas: one on the west coast, one inland that split from the west coast group, and one inland that moved toward the east coast. As discussed previously, the receding of the ice sheets provided room for an initial coastal migration followed by an inland path that opened later. It is difficult to tell when the coastal and inland populations may have split from the samples used, since it may have occurred in Meso-America. The UPGMA tree does favor the split of the two coastal groups occurring before the ice sheets since there is an early separation between the two groups. The Huilliche split from the other Native American groups very early. They also have the highest frequency of the O1v542

mutation. Meltzer (2009) describes small scouting parties that appear to have established different coastal populations, creating a significant founder effect. Perhaps the Huilliche ancestors were one such early coastal group, but they have also been associated with >20% European ancestry, which could skew the data and conclusions (Hunley and Healy 2011).

When the two trees are combined, the new phylogenetic tree mimics the 3-wave migratory trajectory into the Americas that the genetic, linguistic and archaeological evidence suggests in previous research (Greenberg et al. 1986; Reich et al. 2012; Sicoli and Holton 2014; Williams et al. 1985). There is a fourth split near Meso-America that leads to central South American populations that differ slightly from the more coastal populations. The primary branches from the combined tree can be stretched and turned to illustrate their potential as migratory paths (Figure 15). Phylogenetic trees provide telling images of how the original peoples may have moved into and throughout the Americas.


**Figure 15:** Potential migratory paths created from combined NJ/UPGMA tree. Light-grey names are outliers that do not fit migratory paths. Different color paths represent radial expansions from an ancestral Beringian population.

*5.2.4 Mantel Test*

The Mantel test reveals little correlation between genetic distances determined by the K2P model and geographic distances according to longitudinal and latitudinal locations. There are many possible explanations for this. The first and most likely explanation is the imprecision of isolation by distance calculations that assume spatial homogeneity. In other words, when landscape and terrain are not accounted for, it is difficult to substantiate results. To fix this issue, the distances would need to be corrected according to a spatial heterogeneity model like least cost paths (LCP) or circuit theory (Sakaguchi et al. 2010; Sherman et al. 2010; McRae and Shah 2009). This would take extensive knowledge and many hours of rasterizing one and a half continents, which is outside the scope of this project. The second explanation is that there is no geographic correlation with genetics because the multiple migrations (coastal *versus* inland) can lead to groups near each other that separated much earlier. This explanation cannot be qualified without testing the data for heterogeneous landscapes. With an understanding of changing landscapes (especially when the project includes a land bridge that does not exist anymore), varying terrain difficulty, and the potential use of boats, the Mantel Test does not prove to be an acceptable statistic for this study.

**5.3 Gene Flow and Back Migration**

Recurrent gene flow between American and northeast Asian populations has been suggested from genetic and morphological data (Azevedo et al. 2011). The average amount of gene flow required to create the observed frequency of the O1v542 mutation

varies from 5.7-16.5%. The minimum gene flow is between 1.7 and 5.6%, but that is using the population with the highest percentage of O1v542, which also happens to be the population furthest away (Huilliche). The maximum gene flow calculation is the most valuable since it also uses the population that is closest each time. Looking at the maximum value alone, it does not seem likely that gene flow could account for the frequency of the O1v542 allele in all four populations. Schroeder et al. (2007) found that the percent of back migration necessary to account for the 9RA would be 91.9% for the Aleut/Eskimo populations. The combination of the two potential ancestral markers suggests back migration was not the source of the alleles in the Aleut, Chukchi or Koryak populations. It is more likely that the Chukchi and Koryak were founded by back migrations from a source ancestral population.

However, the O1v542 mutation in the Altai could be the result of a back migration or more recent admixture. The O1v542 frequency is just over 1% and only one individual with the mutation was found, so it is reasonable to concede the possibility of the one allele resulting from gene flow. As mentioned previously, individuals in other populations with the O1v542 allele have been observed (Yip 2000, Yip et al. 2006). More Altai samples need to be analyzed in order to validate or contradict the current findings. If the haplotype were found to be consistently present in native Altai, then O1v542 would not be a marker that represents the Beringian source population, but instead it would be a marker that represents the Siberian source population (unless it could then be traced back further). If it were discovered that the O1v542 allele was not found in other Altai samples, then the evidence would point even stronger toward it being an ancestral marker like the 9RA. Further analysis needs to be done, but currently O1v542

appears to be a remnant of a past population that forged across the Bering Strait and pushed around and through the blanketing ice sheets to populate the New World.

## Chapter 6: Conclusion

This study represents the first time the O1v542 mutation has been observed in significant frequencies outside of the Americas. Both Estrada-Mena et al. (2010) and Villanea et al. (2013) put forward a call to examine the presence or absence of the potential ancestral marker in hopes of tracing the migratory and historical connection between Siberian populations and those across North and South America. Through the data collected and analyses performed, more clarity regarding the ancestral lineages of Native Americans is possible.

The Altai, Chukchi, Koryak and Aleut all showed the presence of the O1v542 mutation. They also had higher frequencies of the A1 and B1 markers than any of the other studied populations. Some of the previous researchers indicated a bias toward type O, and the intentional selection of samples with that blood type, which may skew the results (Estrada-Mena et al. 2011). The samples that were screened were all from Meso or South American populations that are known to have O frequencies between 95% and 100%, meaning the data are still reliable but not without critique. Other O haplotypes besides the three primary—O1, O1v and O1v542—appeared sporadically in different populations but never at a high frequency. The exception is O05, which had a high frequency in the Chukchi (8%) and appeared in the Altai (1%), Maya (2%), and Cayapa (1%). Haplotypes O32 and O33 were only seen in the Bolivian Aymara, suggesting recent admixture from European population, which was previously suggested (Hunley and Healy 2011). The Ov7 was only seen in one individual in the Cayapa, which is probably the result of admixture as discussed in the previous chapter.

O1v542 cannot be fully substantiated as an Ancestral Informative Marker (AIM) based on the results since it was traced as far back geographically as this study looked (Altai). The uncertainty stems from inconclusive results because the Altai only had one individual with the marker (1%), which has occurred in other non-Native American populations (Yip 2000, Yip et al. 2006). If the marker can continue to be traced back past the Altai, then it is not an AIM for Native Americans. The upside is even if O1v542 does not prove to be an AIM, it still shows the impact of evolution on the Native American migratory populations because of the shift in allele frequency.

The specific evolutionary force(s) responsible for the increased frequency of the O blood group can be inferred by looking at the frequencies of the O subtypes. Since it is believed no O haplotype holds a selective advantage over any of the other O haplotypes, an increased frequency of one haplotype across populations must be the result of genetic drift. The O1v and O1v542 haplotypes both see substantial frequency increases moving along the potential southerly migratory path, but the O1 and O05 haplogroups do not follow the same design. The increase of O1 and O1v542 follows a similar pattern to the overall increase of the O blood group, suggesting founder effect as the source. Selective events could also have occurred along with the drift, but it is clear founder events played a major role.

If O1v542 is an AIM, then the uniformity of expression of the haplotype across Native Americans and Beringian populations indicates the presence an ancestral population located around Beringia, which spread into the Americas and back toward Siberia. This, along with the 9RA and the M3 Y-chromosome marker, would support the Beringia Incubation Model (BIM). The data from the MDS and phylogenetic trees

support at least 3-waves pulsing into the Americas after a pause in the Beringian area. The first wave traveled along the west coast, passing through the Tlingit population and then splitting into two groups: a.) Parakana and Aymara Chile and b.) Mazahua and Aymara Bolivia. The second-wave started inland and finished along the east coast, connecting the Nahua, Maya, Kayapo, Yanomama, Arara, and eventually stretching back all the way the Huilliche. The third wave leads to the Aleut. Although the Mantel Test shows no significant correlation between genetics and geography, the MDS clusters and the phylogenetic branches connect populations along hypothesized migratory paths. Perhaps a more robust technique for determining geographic distances would lead to a stronger correlation as discussed above.

The gene flow calculations did not reveal back migration from the Americas as being responsible for the O1v542 haplotype in Siberia; instead, they suggest the Chukchi and Koryak may have stemmed from the same ancestral population as all Native Americans. This would add to the 3-wave model, signifying a fourth wave that moved back into Siberia from the incubation point. The migration could have moved all the way back to the Altai region, bringing the O1v542 haplotype with it. This hypothesis would be difficult to test, although the amount of gene flow necessary to account for the frequency of O1v542 in the Altai is reasonably low (~5.3%) to make it feasible. Recent linguistic data also support a connection between the Native American Na-Dene and Siberian Yeniseian that are neighbors of the Altai (Sicoli 2014).

Taken together, the new evidence presented here shows O1v542 is still a valid ancestral informative marker, but further analysis of Altai and other Siberian populations is needed. The evidence also indicates that O1v542 presence in the Chukchi

and Koryak is the result of radial expansion from an ancestral population and not back migration, thus adding to the previously proposed 3-wave model. This theory is supported by Tamm et al. (2007) but opposed by Rubicz et al. (2010). The primary argument of Rubicz et al. (2010) against Kamchatkans stemming from the same ancestral population as Native Americans is a genetic discontinuity with the Aleut population, so they propose back migration as an explanation for the frequency of the 9RA. The Aleuts are a more recent migratory group compared to other Native Americans, so the differences between the Aleut and Kamchatkans is expected, just as differences between the Aleut and Maya are expected. Also, the amount of back migration required to create the frequency of the 9RA seen in the Chukchi and Koryak is ~92%, and according to the results presented here, both the Chukchi and the Koryak are more closely related to the other Native American populations than to the Aleut. The likely explanation is that the Chukchi and Koryak split from the same ancestral source as the Aleut earlier on and have separated genetically since that time. Gene flow certainly played a role across all population throughout the entire process with small movements back and forth, but four major dispersions after the BIM best explain the distributions of the ABO haplotypes observed to date. As more populations are sampled, more archaeological sites are unearthed, and more markers and methods for analysis implemented, the current hypotheses may shift and new connections may be revealed.

## References


Anderson P, Edwards M and Brubaker L (2004). Results and paleoclimatic implications of 35 years of paleoecological research in Alaska. In *The Quaternary period in the United States*, edited by A. Gillespie, S.C. Porter, and B. Atwater, pp. 427-40. Elsevier Science, New York.

Azevedo S, Nocera A, Paschetta C, Castillo L, Gonzalez M, Gonzalez-Jose R (2011). Evaluating Microevolutionary models for the early settlement of the New World: the importance of recurrent gene flow with Asia. *American Journal of Physical Anthropology* 146: 539-552.

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16:37-48.

Barbujani G, Magagni A, Minch E, and Cavalli-Sforza L (1997). An Apportionment of Human DNA Diversity. *Proceedings of the National Academy of Sciences*, USA 90: 4670-4673.

Barjas-Castro ML, Soares MCP, Menezes RC, Carvalho MHM, Costa FF, Saad STO (2003). ABO blood group in Amerindians from Brazilian Amazon. *Hum Biol* 30:220–224.

Bernstein, F. 1931 Die geographische Verteilung der Blutgruppen und ihre anthropologische Bedeutung. *In Comitato Italiano per lo Studio dei Problemi della Populazione,* vol. 3 (ed. C. Ginni), pp. 227–245. Rome, Italy: Instituto Poligrafico dello Stato Liberia.

Bever MR (2006). Too little, too late? The radiocarbon chronology of Alaska and the peopling of the New World. *American Antiquity* 71: 959-620.

Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasaned PK, Claton, HD, Hutz MH, Layrisse Z, Petzyl M, Tsuneto LT, Hill K, Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D and Ruiz-Linares A (2003). Y-chromosome evidence for differing ancient demographic histories in the Americas. *American Journal of Human Genetics* 73: 524-539.

Bradley B, Stanford D (2004). The North Atlantic ice-edge corridor: a possible Palaeolithic route to the New World. *World Archaeology* 36 (4): 459–478.

Campbell L (2011). Review of The Dene-Yeniseian Connection, (ed. James Kari and Ben A. Potter). *IJAL* 77: 445–451.

Crawford MH (1973). The Use of Genetic Markers of the Blood in the Study of the Evolution of Human Populations. In: *Methods and Theories of Anthropological Genetics*. Eds Crawford MH and Workman GL. University of New Mexico Press, Albuquerque. 19-38.

Crawford MH and Workman PL eds. (1973). *Methods and Theories of Anthropological Genetics*. University of New Mexico Press, Albuquerque, NM.

Crawford MH (2007a). *Anthropological Genetics; Theory, Methods and Applications* ed. Crawford MH. Cambridge University Press, New York.

Crawford MH (2007b). Genetic structure of circumpolar populations: a synthesis. *American Journal of Human Biology* 19: 203-217.

Derenko MV, Malyarchuk BA, Grzybowski T, Denosiva G, Dambueva I, Perkova M, Dorzhu C, Luzina F, Lee HK, Vanecek T, Villems R, and Zakharov I (2007). Phylogenetic Analysis of Mitochondrial DNA in Northern Asian Populations. *American Journal of Human Genetics* 81: 1025-1041.

Derenko M, Malyarchuk B, Grzybowski T, Denisova G, Rogalla U, Perkova M, Dambueva I, and Zakharov I (2010). Origin and post-glacial dispersal of mitochondrial DNA haplogroups C and D in northern Asia. *PLoS ONE* 5(12): e15214.

Diamond J (2011). Linguistics: Deep relationships between languages. *Nature* 476: 291–292.

Dillehay TD, Ramirez C, Pino M, Collins MB, Rossen J, Pino-Nacarro JD (2008). Monte Verde: seaweed, food, medicine, and the peopling of South America. *Science* 320: 784-786.

Dulik MC, Zhadanov SI, Osipova LP, Askapuli A, Gau L, Gokcumen O, Rubinstein S, and Schurr TG (2012). Mitochondrial DNA and Y Chromosome Variation Provides Evidence for a Recent Common Ancestry between Native Americans and indigenous Altaians. *American Journal of Human Genetics* 90: 1-18

Dyke AS, Prest VK (1987). Late Wisconsin and Holocene history of the Laurentide Ice Sheet. *Geographic Physique et Quaternaire* 61(2):237-263.

Elias SA (2001). Beringian paleoecology: Results from the 1997 workshop. *Quaternary Science Reviews* 20:7-13.

Estrada-Mena B., Etrada F.J., Ulloa-Arvizu R., Guido M., Mendez R., Coral R., Canto T., Granados J., Rubi-Castellanos R., Rangel-Villalobos H., and Garcia-Carranca A. Blood group O alleles in Native Americans: implications in the peopling of the Americas. *American Journal of Physical Anthropology* 142:85-94.

Excoffier L, Laval G, and Schneider S (2005). Arlequin (version 3.0): An Integrated Software Package for Population Genetics Data Analysis. *Evolutionary Bioinformatics*. 1: 47-50.

Excoffier L and Lischer H (2010). Arlequin Suite ver 3.5: A New Series of Programs to Perform Popualtion Genetics Analyses Under Linux and Windows. *Molecular Ecology Resources*. 10;3: 564-567.

Excoffier L, Smouse PE, and Quattro JM (1992). Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics* 131: 479-491.

Fagundes NJR, Kanitz R, Bonnato SL (2008). A Reevaluation of the Native American mtDNA Genome Diversity and Its Bearing on the Models of Early Colonization of Beringia. *PLoS ONE* 3(9): e3157.

Gagneux P, Varki A (1999). Evolutionary considerations in relating oligosaccharide diversity to biological function. *Glycobiology* 9:747–755.

Goebel T, Waters MR, and O'Rourke DH (2008). The late Pleistocene dispersal of modern humans in the Americas. *Science* 319: 1497-1502.

Greenberg JH, Turner II CG, Zegura SL (1986). The settlement of the Americas: a comparison of the linguistic, dental, and genetic evidence. *Current Anthropology* 27(5): 477-487.

Greenhill SJ, Atkinson QD, Meade A, Gray RD (2010). The shape and tempo of language evolution. *Proc. R. Soc. B* 277: 2443–2450.

Guthrie R (1990). *Frozen fauna of the mammoth steppe: the story of Blue Babe*. University of Chicago Press, Chicago.

Helgason A, Nicholson G, Stefánsson K, and Donnelly P (2003). A Reassessment of Genetic Diversity in Icelanders: Strong Evidence from Multiple Loci for Relative Homogeneity Caused by Genetic Drift. *Annals of Human Genetics*. 67: 281-297.

Hoffecker JF and Elias SA (2007). *Human ecology of Beringia*. Columbia University Press, New York.

Hunley K and Healy M (2011). The impact of founder effects, gene flow, and European admixture on Native America genetic diversity. *American Journal of Physical Anthropology* 146: 530-538.

Jackson LE, Duk-Rodkin A (1996). Quaternaru Geology of the ice-free corridor: glacial controls on the peopling of the New World. In *Prehistoric Mongoloid Dispersals*, edited by T. Akazawa and E.J.E. Szarhmary, pp. 214-227. Oxford University Press, New York.

Kruskal JB (1964a). Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrics* 29: 1-27.

Kruskal JB (1964b). Nonmetric Multidimensional Scaling: A Numeric Method. *Psychometrics* 29: 28-42.

Kruskal JB and Wish M (1978). *Multidimensional Scaling.* Sage Publications, Thousand Oaks, CA.

Lambeck K, Yokoyama Y and Purcell T (2002). Into and out of the Last Glacial Maximum: sea level change during Oxygen Isotope Stage 3 and 2. *Quaternary Science Reviews* 21:343-60.

Landsteiner K (1900). Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt Bakteriologie* 27: 357–62.

Llop E, Henrıquez H, Moraga M, Castro M, Rothhammer F (2006). Brief communication: molecular characterization of O alleles at the ABO locus in Chilean Aymara and Huilliche Indians. *American Journal of Physical Anthropology* 131:535–538.

Malhi RS and Smith DG (2002). Brief communication: Haplogroup X confirmed in prehistoric North America. *American Journal of Physical Anthropology* 119: 84-86.

McRae BH, and Shah VB (2009). Circuitscape User's Guide. ONLINE. The University of California, Santa Barbara. Available at http://www.circuitscape.org.

Meltzer DJ (2009). *First peoples in a new world: colonizing ice age America*. University of California Press, Berkley.

Molnar S (2002). *Human variation.* Upper Saddle River, NJ: Prentice Hall.

Morlan RE, Dyke AS, and McNeely RN (1999). Mapping ancient history, Web page, http://www.geoserv.org.

Morlan RE (2003). Current perspectives on the Pleistocene archeology of eastern Beringia. *Quaternary Research* 60: 123-132.

Nei M (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, USA.

Nei M and Kumar S (2000). *Molecular Evolution and Phylogenetics*. Oxford University Press, USA.

Nei M, and Li WH (1979). Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases. *Proceedings of the National Academy of Sciences*, USA. 76: 5269-5373.

Nichols J (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press. 374 p.

Nicholson G, Smith AV, Jónsson F, Gústafsson Ó, Stefánsson K, and Donnelly P (2002). Assessing Population Differentiation and Isolation from Single-Nucleotide Polymorphism Data. *Journal of the Royal Statistical Society: Series B* 64:695-715.

O'Rourke D (2009). Human migrations: The two roads taken. *Current Biology* 19: R203-R205.

Ogasawara K., Bannai M., Saitou N., Yabe R., Nakata K., Takenaka M., Fujisawa K., Uchikawa M., Ishikawa Y., Juji T., and Tokunaga K (1996). Extensive polymorphism of ABO blood group gene: three major lineages of the alleles for the common ABO phenotypes. *Human Genetics* 97:777-783.

Olsson ML, Santos SEB, Guerreiro JF, Zago MA, Chester MA (1998). Heterogeneity of the O Alleles at the Blood Group ABO Locus in Amerindians. *Vox Sanguinis* 74:46–50.

Peakall R and Smouse PE (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28(19): 2537-2539.

Perego UA, Achilli A, Angerhofer N, Accetturo M, Pala M, Olivieri A, Kashani BH, Ritchie KH, Scozzari R, Hong QP, Myres NM, Salas A, Semino O, Bandelt H, Woodward SR, and Torroni A (2009). Distinctive Paleo-Indian migration routes from Beringia marked by two rare mtDNA haplogroups. *Current Biology* 18:1-8.

Pitblado BL (2011). A tale of two migrations: reconciling recent biological and archaeological evidence for the Pleistocene peopling of the Americas. *Journal of Archaeol Research* 19:327-375.

Reich D, Patterson N, Campbell D, Tandon A, Mazieres S, et al. (2012) Reconstructing Native American population history. *Nature* 488: 370-374.

Rolf FJ (2008). *NTSYSpc: Numerican Taxonomy System*, version 2.2 Exeter Publishing, Ltd. Setauket, NY.

Roubinet F, Kermarrec N, Despiau S, Apoil P-A, Dugoujon JM, Blancher A (2001). Molecular polymorphism of O alleles in five populations of different ethnic origins. *Immunogenetics* 53:95–104.

Rubicz R, Melton PE, Spitsyn V, Sun G, Deka R, and Crawford MH (2010). Genetic structure of native circumpolar populations based on autosomal, mitochondrial, and T-chromosome DNA markers. *American Journal of Physical Anthropology* 143: 62-74.

Ruhlen M (1998). The origin of the Na-Dene. PNAS 95: 13994–13996.

Sakaguchi T, Morin J, Dickie R (2010). Defensibility of large prehistoric sites in Mid-Fraser region on the Canadian plateau. *Journal of Archaeological Science* 37:1171-1185.

Schroeder KB, Schurr TG, Long JC, Rosenburg NA, Crawford MH, Tarskaia LA, Osipova LP, Zhadanov SI, and Smith DG (2007). A private allele ubiquitous in the Americas. *Biology Letters* 3: 218-223.

Schroeder KB, Jakobsson M, Crawford MH, Schurr TG, Boca SM, Conrad DF, Tito RY, Osipova LP, Tarskaia LA, Zhadanov SI, Wall JD, Pritchard JK, Malhi RS, Smith DG, Rosenberg NA (2009). Haplotypic background of a private allele at high frequency in the Americas. *Mol Biol Evol* 26(5): 995-1016.

Schurr TG (2004). The peopling of the New World: Perspectives from molecular anthropology. *Annual Review of Anthropology* 33: 551-583.

Sherman RJ, Balkansky AK, Spencer CS, Nicholis B (2010). Expansionary dynamics of the nascent Monte Alban state. *Journal of Anthropological Archaeology* 29:278-301.

Sicoli MA, Holton G (2014). Linguistic phylogenies support back-migration from Beringia to Asia. *PLoS ONE* 9(3): e91722.

Sokal R and Michener C (1958). A statistical method for evaluating systematic relationships. University of Kansas Science Bulletin 38: 1409–1438.

Stone AC and Stoneking M (1999). Analysis of Ancient DNA from a Prehistoric Amerindian Cemetery. Philosophical Transactions of the Royal Society of London. *Biological Sciences*. 354: 153-159.

Swerdlow DL, Mintz ED, Rodriguez M, Tejada E, Ocampo C, Espejo L, Barrett TJ, Petzelt J, Bean NH, Seminario L, Tauxe RV (1994). Severe life-threatening cholera associated with blood group O in Peru: implications for the Latin American epidemic. *J Infect Dis* 170:468–472.

Szathmary EJE (1979). Blood groups of Siberians, Eskimos, and subarctic and Northwest Coast Indians: the problem of origins and genetic relationships. In: Laughlin WS, Harper AB, editors. *The First Americans: origins, affinities, and adaptations*. New York: G. Fischer. p 185–209.

Tajima F (1983). Evolutionary Relationship of DNA Sequences in Finite Popualtions. *Genetics*. 123: 585-595.

Tajima F (1989). Statistical Method for Testing the Neutral Mutation Hypothesis DNA Polymorphism. *Genetics*. 123: 585-595.

Tamm E, Kivisild T, Reidla M, Metspalu M, Smith DG, Mulligan CJ, Bravi CM, Rickards O, Martinez-Labarga C, Khusnutdinova EK, Fedorova SA, Golubenko MV, Stepanov VA, Gubina MA, Zhadanov SI, Ossipova LP, Damba L, Voevoda MI, Dipierri JE, Villems R and Malhi RS (2007). Beringian standstill and spread of Native American founders. *PLoS ONE* September 2007: e829.

Tamura K, Dudley J, Nei M, and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 *Molecular Biology and Evolution* 24: 1596-1599.

Tamura K, Peterson D, Peterson N, Stecher G, Masatoshi N, and Kumar S (2011). MEGA5: Analysis Using Maximum Likelihood, Evolutionary Distance and Maximum Parsimony. *Molecular Biology and evolution*. 28(10): 2731-2739.

Torgerson WS (1952). Multidimensional Scaling, 1: theory and method. *Psychometrics* 17: 401-419.

Vajda E (2010) Yeniseian, Na-Dene, and historical linguistics. *APUA New Series* 5: 100–118.

Villanea FA, Bolnick DA, Monroe C, Worl R, Cambra R, Leventhal A, and Kemp BM (2013). Brief Communication: Evolution of a specific O allele (O1vG542A) supports unique ancestry of Native Americans. *American Journal of Physical Anthropology* 151:649-657.

Waguespack NM (2007). Why we're still arguing about the Pleistocene occupation of the Americas. *Evolutionary Anthropology* 16: 63-74.

Wang S, Lewis Jr. CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A (2007). Genetic variation and population structure in Native Americans. *PLoS* 3: 2049-2067.

Williams RC, Steinberg AG, Gershowitz H, Bennett PH, Knowler WC, et al. (1985) GM allotypes in Native Americans: Evidence for three distinct migrations across the Bering land bridge. *American Journal of Physical Anthropology* 66: 1–19.

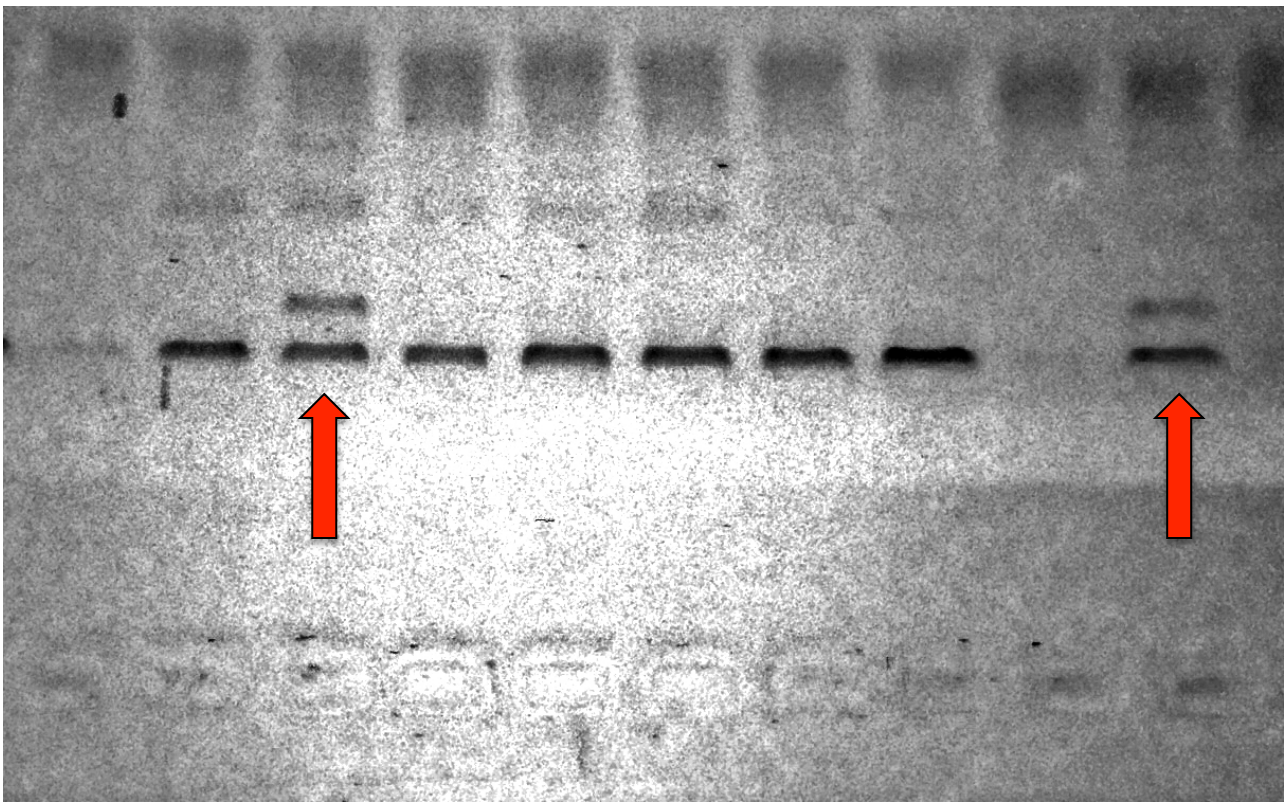
Yip SP (2000). Single-tube multiplex PCR-SSCP analysis distinguishes 7 common ABO alleles and readily identifies new alleles. *Blood* 95:1487–1492.

Yip SP (2002). Sequence variation at the human ABO locus. *Ann Hum Genet* 66:1–27.

Yip SP, Choi PS, Lee SY, Leung KH, El-Zawahri Mokhtar M, Luqmani Yunus A (2006). ABO blood group in Kuwaitis: detailed allele frequency distribution and identification of novel alleles. *Transfusion* 46:773-779.

## Appendix

A. Sample electrophoresis gel run after restricting exon 7 with the NheI enzyme. In the image, two different samples have multiple bands, indicating heterozygosity at the 542 position. No individuals were homozygous, so all O1v542 individuals were confirmed by the presence of two bands. The two heterozygote samples are indicted by arrows.



B. Raw sequence data collected from all four populations. C = Chukchi, A= Altai, E = Koryak, and Aleut = Aleut. The two haplotype alleles for each individual are listed in columns 2 and 3.

| Sample | ABO | ABO | Exon 6 | | Exon 7 | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Allele | Allele | 261 | 297 | 467 | 498 | 526 | 538 | 542 | 646 | 657 | 681 | 703 | 771 | 796 | 803 | 829 | 930 |
| Reference | A1 | A1 | G | A | C | C | C | C | G | T | C | G | G | C | C | G | G | G |
| C7 | A102 | A102 | ** | ** | TT | ** | | ** | | | | | | | | | | |
| C11 | O05 | A102 | *- | G* | *T | ** | | ** | | | | | | | | | | |
| C15 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| C16 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C17 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| C27 | O05 | O1 | -- | *G | | ** | | ** | | ** | | ** | | ** | | | ** | |
| C28 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C29 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| C30 | O1v542 | O1 | -- | G* | | ** | | ** | *A | A* | | A* | | T* | | | A* | |
| C32 | A102 | A102 | ** | ** | TT | ** | | ** | | ** | | ** | | ** | | | ** | |
| C38 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| C39 | O05 | A102 | *- | G* | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| C42 | O05 | A102 | *- | G* | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| C45 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C48 | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A |
| C51 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| C52 | A101 | O1 | *- | G* | | ** | | ** | | | | | | | | | | |
| C53 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| C54 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C56 | O05 | A101 | -- | *G | | ** | | ** | | ** | | ** | | ** | | | ** | |
| C57 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C60 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| C61 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C63 | O05 | O1 | -- | G* | | ** | | ** | | ** | | ** | | ** | | | ** | |
| C64 | O1v542 | O1 | -- | G* | | ** | | ** | *A | A* | | A* | | T* | | | A* | |
| C65 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C66 | O05 | O1 | -- | G* | | ** | | ** | | ** | | ** | | ** | | | ** | |
| C67 | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A |
| C68 | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A |
| C69 | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A |
| C70 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| C71 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C74 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C76 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C77 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| C78 | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A |
| C79 | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| C80 | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| C82 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| C83 | O1 | B101 | *- | G* | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| C84 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| C85 | O1 | B101 | *- | G* | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A1 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| A2 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A3 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| A4 | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| A8 | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| A9 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A11 | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A15 | O1 | B101 | *- | G* | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A16 | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A17 | A101 | A102 | ** | ** | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| A19 | B101 | B101 | ** | GG | | ** | GG | ** | | ** | TT | ** | AA | ** | AA | CC | ** | AA |
| A20 | O1v542 | A102 | *- | G* | *T | ** | | ** | A* | A* | | A* | | T* | | | A* | |
| A23 | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| A28 | O1 | B101 | *- | G* | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A31 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| A32 | O1 | O1 | -- | ** | | ** | | ** | | ** | | ** | | ** | | | ** | |
| A38 | O05 | A102 | *- | G* | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| A40 | O1 | B101 | *- | GG | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A41 | A102 | A102 | ** | ** | TT | ** | | ** | | ** | | ** | | ** | | | ** | |
| A42 | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A |
| A43 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| A44 | O1 | B101 | *- | G* | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A45 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| A46 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A47 | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A |
| A50 | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A51 | O1 | A101 | *- | ** | | ** | | ** | | ** | | ** | | ** | | | ** | |
| A52 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A54 | O1 | B101 | *- | G* | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A55 | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | |
| A56 | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A61 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A64 | O1 | B101 | *- | G* | | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A67 | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| A68 | O1v | A101 | *- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A70 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A71 | O1v | O1 | -- | GG | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A73 | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| A77 | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | ** | | ** | | | ** | |
| A80 | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | |
| A82 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| A85 | O1v | O1v | -- | GG | | ** | | ** | A* | AA | | AA | | TT | | | AA | |
| Aleut300 | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | ** | *A | ** | *A | *C | ** | *A |
| Aleut301 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| Aleut304 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |
| Aleut305 | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Aleut307** | O1 | O1 | -- | ** | | ** | | ** | | ** | | ** | | ** | | | ** | | |
| **Aleut308** | A102 | A102 | -- | ** | TT | ** | | ** | | ** | | ** | | ** | | | ** | | |
| **Aleut310** | O1 | O1 | -- | ** | | ** | | ** | | ** | | ** | | ** | | | ** | | |
| **Aleut311** | O1v542 | A101 | *- | G* | | ** | | ** | A* | A* | | A* | | T* | | | A* | | |
| **Aleut315** | O1 | A101 | *- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut318** | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **Aleut320** | O1 | A101 | *- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut321** | O1 | A101 | *- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut324** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut333** | O1 | B101 | *- | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **Aleut339x** | A101 | A102 | ** | ** | *T | ** | | ** | | ** | | | | | | | | | |
| **Aleut339y** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut340** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **Aleut341** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **Aleut355** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **Aleut357** | O1 | B101 | *- | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **Aleut373** | A102 | A102 | ** | ** | TT | ** | | ** | | ** | | | | | | | | | |
| **Aleut375** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **Aleut376** | O1 | A101 | *- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut379** | O1v | A101 | *- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **Aleut386** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut389** | A102 | A102 | ** | ** | TT | ** | | ** | | ** | | | | | | | | | |
| **Aleut390** | O1v542 | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A | |
| **Aleut391** | B101 | A101 | ** | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **Aleut397** | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | | | | | | | | |
| **Aleut405** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut407** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut408** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **Aleut409** | O1 | A101 | *- | ** | *T | ** | | ** | | ** | | | | | | | | | |
| **Aleut419** | O1 | B101 | *- | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **Aleut420** | O1 | B101 | *- | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **Aleut421y** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **Aleut440** | B101 | A101 | ** | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **E5** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E15** | O1 | A102 | *- | ** | *T | ** | | ** | | ** | | | | | | | | | |
| **E17** | O1v | B101 | *- | GG | | ** | *G | ** | | A* | *T | A* | *A | T* | *A | *C | A* | *A | |
| **E27** | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E35** | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | | |
| **E41** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E45** | O1v542 | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E50** | A102 | O1 | *- | ** | *T | ** | | ** | | ** | | | | | | | | | |
| **E54** | A102 | B101 | ** | G* | *T | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **E58** | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | | |
| **E59** | O1 | B101 | *- | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **E63** | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | | |
| **E66** | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | | |
| **E67** | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | | |
| **E77** | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E78** | O1v | A102 | *- | G* | *T | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E85** | O1v | A101 | *- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E95** | O1v542 | A101 | *- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E102** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E114** | O1v | O1 | -- | G* | | ** | | ** | | A* | | A* | | T* | | | A* | | |
| **E131** | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | | |
| **E137** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **E145** | O1 | O1 | -- | ** | | ** | | ** | | ** | | | | | | | | | |
| **E151** | O1 | B101 | *- | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |
| **E164** | O1v | O1v | -- | GG | | ** | | ** | | AA | | AA | | TT | | | AA | | |
| **E165** | O1 | B101 | *- | *G | | ** | *G | ** | | ** | *T | | *A | | *A | *C | | *A | |